프로젝트 개요

이 프로젝트는 데이터 증강(data augmentation)을 기반으로 디지털 미디어 콘텐츠를 생성하는 시스템입니다. 특히 이미지 전처리, 증강 및 Stable Diffusion 모델을 활용한 이미지 생성 및 미세 조정을 포함합니다.

주요 기능 및 역할

- 데이터 전처리: 이미지 크기 조정, tif 파일을 png로 변환, padding 처리
- 데이터 증강: 좌우 반전, 상하 반전, 흑백 변환, 채도 및 명도 조절, 회전 변환
- 모델 로딩: Stable Diffusion v1-5 모델을 GPU(CUDA) 환경에서 불러와 사용
- 미세 조정(Fine tuning): PyTorch, torchvision 기반으로 LoRA(Low-Rank Adaptation) 기법을 활용하여 image generation 모델 학습
- 학습 환경 설정: 혼합 정밀도(fp16) 사용, 학습 스케줄러 및 옵티마이저 구현
- 모델 저장 및 로딩: 학습 완료 후 모델을 저장하고, 재사용을 위한 모델 불러오기 기능
- 이미지 생성 테스트: 임의 프롬프트로 생성 이미지 테스트 및 파일 저장 기능
- 평가 함수: FID(Frechet Inception Distance)를 통한 생성 이미지 품질 평가 구현

사용 기술

- Python 3.x
- PyTorch, torchvision, torchaudio
- Diffusers, Huggingface Hub, transformers
- PIL, datasets
- CUDA, mixed precision, LoRA
- Google Colab 환경

시스템 구성도 (요약)

1. 데이터 전처리 및 증강
2. 모델 로딩 및 미세 조정
3. 학습 루프 및 평가
4. 모델 저장 및 재사용
5. 이미지 생성 및 결과 저장

담당 역할

- 데이터 증강 및 전처리 파이프라인 설계 및 구현
- Stable Diffusion 기반 이미지 생성 모델의 미세 조정 및 최적화
- PyTorch 프레임워크를 활용한 학습 루프 개발, 혼합 정밀도 및 메모리 최적화 적용
- LoRA 기법을 적용한 모델 압축 및 효율적 학습 구현
- FID를 통한 생성 결과 평가 및 성능 향상 작업
- Google Colab과 연동한 원격 학습 환경 구축 및 관리
- 이미지 자동 저장, 버전 관리 기능 추가

구현 성과

- 다양한 증강 기법을 실시간으로 적용하여 데이터셋 다양성 및 모델 강건성 향상
- LoRA 적용으로 학습 파라미터 효율성 증대 및 메모리 사용량 감소
- Stable Diffusion 기반 이미지 생성 성능 및 품질 개선
- 자동화된 이미지 저장/관리 및 평가 체계 구축으로 연구 반복성과 재현성 확보

주요 사용 기술 및 도구

- Python, PyTorch, torchvision, diffusers, huggingface transformers
- CUDA GPU 환경, mixed precision(training speedup)
- Google Colab, Google Drive 연동
- LoRA(Low-Rank Adaptation) for parameter-efficient tuning
- FID 평가 지표 활용

【서지사항】

【서류명】 특허출원서

【참조번호】 PN159551KR

【출원구분】 특허출원

【출원인】

【성명】 두지언

【특허고객번호】 4-2023-000001-9

【대리인】

【명칭】 리앤목 특허법인

【대리인번호】 9-2005-100002-8

【지정된변리사】 이영필,이해영,민관호

【발명의 국문명칭】 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성 시스템

【발명의 영문명칭】 A system for generating digital media contents based on data augmentation

【발명자】

【성명】 두지언

【특허고객번호】 4-2023-000001-9

【발명자】

【성명】 이정민

【성명의 영문표기】 LEE, Jung Min

【국적】 KR

【주민등록번호】 900723-0XXXXXX

【우편번호】 35224

【주소】 대전광역시 서구 계룡로319번길 13 (월평동)

【거주국】 KR

【발명자】

【성명】 이세현

【성명의 영문표기】 LEE, Se Hyeon

【국적】 KR

【주민등록번호】 980402-0XXXXXX

【우편번호】 50917

【주소】 경상남도 김해시 가락로125번길 40, 301호 (봉황동)

【거주국】 KR

【출원언어】 국어

【심사청구】 청구

【취지】 위와 같이 특허청장에게 제출합니다.

대리인 리앤목 특허법인 (서명 또는 인)

【수수료】

【출원료】 0 면 46,000 원

【가산출원료】 143 면 0 원

【우선권주장료】 0 건 0 원

【심사청구료】 28 항 1,594,000 원

【합계】 1,640,000원

【감면사유】 개인(70%감면)[1]

【감면후 수수료】 492,000 원

【수수료 자동납부번호】 3010358854671

【첨부서류】 1.기타첨부서류[위임장]_1통

【발명의 설명】

【발명의 명칭】

데이터 증강에 기반한 디지털 미디어 컨텐츠 생성 시스템{A system for generating digital media contents based on data augmentation}

【기술분야】

【0001】 본 발명은 데이터 증강에 기반한 미디어 컨텐츠 생성 시스템에 관한 것이다.

【발명의 배경이 되는 기술】

【0002】 AI 기술의 발전에 따라 text-to-image, image-to-text와 같은 멀티 모달의 AI 모델로부터 서로 다른 이종 모달의 텍스트, 이미지 등의 미디어를 생성할 수 있는 생성형 AI 모델(generative AI model)이 개발되고 있으며, 이러한 생성형 AI 모델은 의료, 금융, 게이밍, 마케팅, 패션과 같은 다양한 분야에서의 활용이 모색되고 있는 실정이며, 최근에는 공간 디자인과 같은 아트 분야의 대중화를 위하여 자신만의 취향이 반영된 아트 전시품을 스스로 구현할 수 있는 AI 기술의 개발이 요구되고 있다.

【발명의 내용】

【해결하고자 하는 과제】

【0003】 본 발명의 일 실시형태는 원본 데이터의 희귀성에도 불구하고 생성형 AI 모델의 파인-튜닝을 위한 충분한 학습 데이터를 확보하고 확보된 파인-튜닝

을 위한 학습 데이터로부터 과적합(over-fitting)이 없이 일반화 능력이 향상되면서, 생성 대상 이미지에 표현될 주제 또는 객체 등에 관한 컨텐츠 정보를 포함하는 이미지 또는 텍스트의 컨텐츠 데이터와, 생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터를 입력으로 하여, 컨텐츠 데이터에 표현된 생성 대상 이미지의 주제 또는 객체 등에 관한 컨텐츠가 스타일 데이터로부터 추출된 스타일로 표현되는 합성 데이터를 생성할 수 있는, 생성형 AI 네트워크를 포함한다.

【과제의 해결 수단】

【0004】 상기와 같은 과제 및 그 밖의 과제를 해결하기 위하여, 본 발명의 일 실시형태에 따른 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성 시스템은,

【0005】 원본 데이터로부터 원본 데이터의 스타일이 전이된 가상 데이터를 생성하기 위한 것으로, 스타일 전이된 이미지를 생성하도록 training된 제1 디지털 미디어 컨텐츠 생성 모델; 및

【0006】 상기 원본 데이터와 상기 제1 디지털 미디어 컨텐츠 생성 모델로부터 생성된 가상 데이터를 포함하는 동일 유사 스타일로 표현된 학습 데이터로부터, 스타일 전이된 이미지를 생성하도록 training된 모델을 대상으로 파인-튜닝 시키기 위한 제2 디지털 미디어 컨텐츠 생성 모델;을 포함하고,

【0007】 상기 제2 디지털 미디어 컨텐츠 생성 모델은 학습 또는 파인-튜닝을 위하여 loss function의 산출을 위한 loss computing 네트워크를 포함하여 loss computing 프로세스를 구현하되,

【0008】상기 제1 디지털 미디어 컨텐츠 생성 모델은 상기 loss computing 네트워크를 포함하지 않거나 또는 loss computing 프로세스를 구현하지 않을 수 있다.

【0009】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은, 상기 제1 디지털 미디어 컨텐츠 생성 모델의 네트워크를 공유할 수 있다.

【0010】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은,

【0011】상기 제1 디지털 미디어 컨텐츠 생성 모델을 포함하고,

【0012】상기 loss computing 네트워크를 더 포함할 수 있다.

【0013】예를 들어, 상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

【0014】생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터(y)로서 원본 데이터 또는 가상 데이터와, 생성 대상 이미지에 표현될 컨텐츠 정보를 포함하는 컨텐츠 데이터(x)를 서로 다른 입력으로 하여,

【0015】컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 instant normalization된 컨텐츠 데이터(t)로부터 Generator G(g)를 통하여 이미지 또는 feature를 생성할 수 있다.

【0016】예를 들어, 상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

【0017】상기 스타일 데이터(y)로부터 추출된 feature statistics로부터 추출된 스케일링(scaling) 파라메타 및 바이어스(bias) 파라메타를 적용한 AdaIN(Adaptive Instant Normalization)을 통하여 instant normalization된 컨텐츠 데이터(t)를 생성할 수 있다.

【0018】예를 들어, 상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

【0019】컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 이미지 또는 feature를 출력하기 위한 Generator G(g)로서, convolution layer 또는 업-스케일링을 위한 Upsample 네트워크를 포함할 수 있다.

【0020】예를 들어, 상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

【0021】저해상도의 feature 또는 feature map으로부터 고해상도의 이미지를 향하여 progressive 하게 업-스케일링으로 전개되는 멀티-스케일을 형성하는 각각의 스케일에서, 상기 Generator G(g)로부터 이미지 또는 feature의 생성이 구현될 수 있다.

【0022】예를 들어, 상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

【0023】 레이턴트 표현(z)을 입력으로 하여 레이턴트 표현(z)의 서로 다른 차원 사이의 disentanglement를 위한 다수의 FC(Fully Connected) layer의 적층과, 상기 다수의 FC layer의 적층으로부터의 출력과 상기 원본 데이터 또는 가상 데이터를 서로 다른 입력으로 하여, 상기 Generator G(g)를 포함하는 Synthesis network로 주입되는 스타일 정보를 추출하기 위한 intermediate 레이턴트 표현(w)을 생성하기 위한 Mapping network를 포함할 수 있다.

【0024】 예를 들어, 상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

【0025】 상기 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature를 추출하기 위한 인코더(Encoder); 및

【0026】 상기 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 이미지(g(t)) 또는 feature(g(t))를 출력하기 위한 Generator G(g)로서, AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t)로부터 차원 확장을 통하여 이미지를 생성하기 위한 디코더(Decoder);를 포함할 수 있다.

【0027】 예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은,

【0028】생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터(y)로서 원본 데이터와 생성 대상 이미지에 표현될 컨텐츠 정보를 포함하는 컨텐츠 데이터(x)를 서로 다른 입력으로 하여,

【0029】컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)을 통하여 instant normalization된 컨텐츠 데이터(t)로부터 Generator G(g)에서 생성된 이미지(g(t)) 또는 feature(g(t))를 입력으로 하는 loss computing 네트워크의 출력(f(g(t))과, AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t) 사이의 contents loss(Lc)와, 상기 loss computing 네트워크의 출력(f(g(t))과 스타일 데이터(y) 사이의 style loss(Ls)를 취합한 loss function을 산출할 수 있다.

【0030】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은,

【0031】상기 loss computing 네트워크의 출력(f(g(t))과 AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t) 사이의 contents loss(Lc)로서, 상기 loss computing 네트워크의 출력(f(g(t))과 AdaIN으로 instant normalization된 컨텐츠 데이터(t)의 feature 또는 feature map 사이에서 이하와 같은 픽셀 단위의 거리(L2 norm)에 기반한 contents loss(Lc)를 산출할 수 있다.

【0032】

【0033】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은,

【0034】상기 loss computing 네트워크의 출력(Φ(g(t))로부터 추출된 feature statistics로서 mean(평균) μ와 standard deviation(표준편차) σ와, 스타일 데이터(y)로부터 추출된 feature statistics로서 mean(평균) μ와 standard deviation(표준편차) σ 사이의 거리(L2 norm)에 기반한 이하와 같은 style loss(Ls)를 산출할 수 있다.

【0035】

【0036】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은, 상기 contents loss(Lc) 및 style loss(Ls)가 스케일 팩터(λ)를 적용하여 합산된 이하와 같은 loss function을 산출할 수 있다.

【0037】

【0038】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은, loss computing 네트워크로부터 산출된 loss function을 최소화시키도록 학습될 수 있다.

【0039】예를 들어, 상기 loss function을 산출하기 위한 loss computing 네트워크는, Generator G(g)의 역-프로세스를 구현하면서,

【0040】Generator G(g)로부터 생성된 이미지(g(t)) 또는 feature(g(t))에 대해 역-프로세스를 적용하여, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)로 복원되는지에 관한 contents loss(Lc); 및

【0041】Generator G(g)로부터 생성된 이미지(g(t)) 또는 feature(g(t))에 대해 역-프로세스를 적용하여, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics를 추종하는지에 관한 style loss(Ls)를 산출할 수 있다.

【0042】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서, 상기 Generator G(g)는 학습 대상에 해당되되, 상기 loss computing 네트워크는 학습 대상에 해당되지 않을 수 있다.

【0043】예를 들어, 상기 Generator G(g)는, 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)을 통하여 instant normalization된 컨텐츠 데이터(t)로부터 차원 확장을 통하여 이미지(g(t)) 또는 feature(g(t))를 생성하기 위한 디코더(Decoder)를 포함하고,

【0044】상기 loss computing 네트워크는, 상기 디코더(Decoder)의 역-프로세스로서 디코더(Decoder)로부터 생성된 이미지(g(t)) 또는 feature(g(t))로부터 차원 축소된 feature를 추출하기 위한 인코더(Encoder)를 포함할 수 있다.

【0045】예를 들어, 상기 제2 디지털 미디어 컨텐츠 생성 모델은,

【0046】상기 컨텐츠 데이터 및 스타일 데이터로부터 feature를 추출하기 위한 제1 인코더를 더 포함하고,

【0047】상기 loss computing 네트워크는 상기 제1 인코더와 동일한 네트워크로 구현될 수 있다.

【0048】예를 들어, 상기 제1 인코더 및 loss computing 네트워크는 CNN(Convolution Neural Network) 계열의 VGG net 기반의 Autoencoder로서 사전에 학습된 VGG net의 convolution layer를 포함할 수 있다.

【0049】예를 들어, 상기 제1 디지털 미디어 컨텐츠 모델 또는 제2 디지털 미디어 컨텐츠 모델은,

【0050】원본 이미지로부터 시간 스텝을 전진시키면서 노이즈 스케줄(noise schedule)로부터 정의된 노이즈를 추가하여 점진적으로 원본 이미지의 패턴을 붕괴시키는 노이징 프로세스(noising process)의 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복원되도록 노이지한 이미지로부터 상대적으로 원본 이미지에 가까운 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현하면서 이미지를 생성할 수 있다.

【0051】예를 들어, 상기 제1 디지털 미디어 컨텐츠 모델과 제2 디지털 미디어 컨텐츠 모델은,

【0052】노이즈를 추가하면서 원본 이미지를 붕괴시키는 노이징 프로세스(noising process)의 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복

원되도록 노이지한 이미지로부터 상대적으로 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현하기 위한 네트워크를 공유할 수 있다.

【0053】 예를 들어, 상기 제1 디지털 미디어 컨텐츠 모델은,

【0054】 이미지 생성을 위한 denoising process를 구현하기 위한 네트워크를 포함하여 denoising process는 수행하되,

【0055】 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 구현하기 위한 네트워크는 포함하지 않거나 또는 학습 또는 파인-튜닝을 위한 forward process로서 noising process는 수행하지 않을 수 있다.

【0056】 예를 들어, 상기 제2 디지털 미디어 컨텐츠 모델은,

【0057】 이미지 생성을 위한 denoising process를 구현하기 위한 네트워크 및 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 구현하기 위한 네트워크를 모두 포함하여, 이미지 생성을 위한 denoising process와 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 모두 수행하면서,

【0058】 상기 denoising process에서 현재의 시간 스텝에서의 이미지(Xt)로부터 덜 노이지한 이전 시간 스텝(Xt-1)에서 추가된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균을 예측하고, noising process에서 산출된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균과의 오차를 최소화 시키도록 상기 denoising process를 학습할 수 있다.

【0059】 예를 들어, 상기 제1 디지털 컨텐츠 생성 모델은,

【0060】이미지 생성 조건으로 이미지 conditioning 또는 텍스트 conditioning의 주입을 위한 인코더를 포함할 수 있다.

【0061】예를 들어, 상기 인코더는,

【0062】상기 이미지 생성 조건으로, 상기 원본 데이터를 이미지 conditioning으로 주입하기 위한 이미지 인코더; 및

【0063】상기 이미지 생성 조건으로, 상기 원본 데이터의 변형 조건 또는 가상 데이터로 표현될 컨텐츠 정보를 텍스트 conditioning으로 주입하기 위한 텍스트 인코더;를 포함할 수 있다.

【0064】예를 들어, 상기 이미지 인코더 및 텍스트 인코더는 각각,

【0065】멀티-모달의 CLIP(contrastive language image pre-training)을 통하여 이미지-텍스트의 멀티-모달(multi modal)의 임베딩 공간을 학습하여,

【0066】이미지 conditioning으로 주입된 원본 데이터를 클립 이미지 임베딩(CLIP image embedding)으로 인코딩하고,

【0067】텍스트 conditioning으로 주입된 원본 데이터의 변형 조건 또는 가상 데이터로 표현될 컨텐츠 정보를 클립 텍스트 임베딩(CLIP text embedding)으로 인코딩할 수 있다.

【0068】예를 들어, 상기 제1 디지털 미디어 컨텐츠 모델은,

【0069】노이즈를 추가하면서 원본 이미지를 붕괴시키는 노이징 프로세스(noising process)의 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복

원되도록 노이지한 이미지로부터 상대적으로 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현하면서,

【0070】 상기 이미지 생성 조건을 주입하기 위한 인코더를 포함하여, 이미지 conditioning 또는 텍스트 conditioning을 이미지 임베딩 또는 텍스트 임베딩으로 인코딩하여 생성 조건으로 주입하기 위한 임베딩 또는 인코딩을 구현할 수 있다.

【0071】 예를 들어, 상기 제2 디지털 미디어 컨텐츠 모델은,

【0072】 이미지 생성을 위한 denoising process를 구현하기 위한 네트워크 및 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 구현하기 위한 네트워크를 모두 포함하여, 이미지 생성을 위한 denoising process와 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 모두 수행하면서,

【0073】 상기 denoising process에서 현재의 시간 스텝에서의 이미지(Xt)로부터 덜 노이지한 이전 시간 스텝(Xt-1)에서 추가된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균을 예측하고, noising process에서 산출된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균과의 오차를 최소화 시키도록 상기 denoising process를 학습하되,

【0074】 상기 이미지 생성 조건을 주입하기 위한 인코더를 포함하지 않거나 또는 이미지 conditioning 또는 텍스트 conditioning을 이미지 임베딩 또는 텍스트 임베딩으로 인코딩하여 생성 조건으로 주입하기 위한 임베딩 또는 인코딩을 구현하지 않을 수 있다.

【발명의 효과】

【0075】본 발명에 의하면, 원본 데이터의 희귀성에도 불구하고 생성형 AI 모델의 파인-튜닝을 위한 충분한 학습 데이터를 확보하고 확보된 파인-튜닝을 위한 학습 데이터로부터 과적합(over-fitting)이 없이 일반화 능력이 향상되면서, 생성 대상 이미지에 표현될 주제 또는 객체 등에 관한 컨텐츠 정보를 포함하는 이미지 또는 텍스트의 컨텐츠 데이터와, 생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터를 입력으로 하여, 컨텐츠 데이터에 표현된 생성 대상 이미지의 주제 또는 객체 등에 관한 컨텐츠가 스타일 데이터로부터 추출된 스타일로 표현되는 합성 데이터를 생성할 수 있는, 생성형 AI 네트워크가 구현될 수 있다.

【도면의 간단한 설명】

【0076】도 1에는 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 원본 데이터(스타일 데이터, 이미지 1)와 컨텐츠 데이터(이미지 2) 또는 원본 데이터의 변형 조건이나 컨텐츠 정보(텍스트)를 입력으로 하여, 가상 데이터(이미지 3, 전이된 스타일로 컨텐츠가 표현된 가상 이미지)를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델(생성형 AI 네트워크)을 설명하기 위한 도면이 도시되어 있다.

도 2에는 본 발명의 일 실시형태에서, 원본 데이터(스타일 데이터, 이미지 4) 및 가상 데이터(스타일 데이터, 이미지 4)를 파인-튜닝의 학습 데이터로 하여, 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 모델(생성형 AI 모델)을 설명하기 위

한 도면으로, 원본 데이터(스타일 데이터, 이미지 4) 또는 가상 데이터(스타일 데이터, 이미지 4)와 컨텐츠 데이터(이미지 5)를 입력으로 하여, 전이된 스타일로 컨텐츠가 표현된 합성 데이터를 생성하도록 파인-튜닝을 구현하는 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

도 3a 및 도 3b에는 본 발명의 일 실시형태에서, 원본 데이터의 변형(transformation)을 통한 가상 데이터의 생성 또는 데이터 증강을 설명하기 위한 도면으로, 각각 (a) Flip, (b) Gray scale, (c) Saturation, (d) Brightness, (e) Rotation의 Img2Img transformation을 설명하기 위한 도면들이 도시되어 있다.

도 4에는 디퓨전의 forward process로서, 특정한 패턴의 선명한 원본 이미지(swiss roll)에 시간의 스텝에 따라 점진적으로 노이즈 스케줄에 따른 노이즈를 추가하면서 원본 이미지의 특정한 패턴이 붕괴된 완전한 가우시안 노이즈(isotropic Gaussian noise)를 생성하는 noising process를 설명하기 위한 도면이 도시되어 있다.

도 5에는 디퓨전의 reverse process로서, 완전한 가우시안(isotropic Gaussian noise)로부터 예측된 노이즈를 제거하면서 시간의 스텝에 따라 덜 노이지한 이미지를 생성하면서, 원본 이미지의 패턴을 복원하는 이미지 생성을 설명하기 위한 도면으로, denoising process를 설명하기 위한 도면이 도시되어 있다.

도 6에는 노이즈가 추가된 이미지와, 해당되는 이미지의 시간 스텝을 입력으로 하여, 현재 시간 스텝에서 추가된 노이즈를 예측함으로써, 현재 시간 스텝에서의 노이지한 이미지로부터 상대적으로 덜 노이지한 이미지의 확률 분포를 예측하기

위한 U-net의 아키텍처를 설명하기 위한 도면이 도시되어 있다.

도 7에는 본 발명의 일 실시형태에서 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 화소 공간 보다 저차원의 레이턴트 공간(latent space) 상에서 디퓨전이 구현되는 레이턴트 디퓨전 모델(LDM, Latent Diffusion Model 또는 stable Diffusion Model)로 구현된 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 원본 데이터(y)를 이미지 conditioning으로 주입하기 위한 이미지 인코더로부터 출력되는 이미지 생성을 가이드 하기 위한 생성 조건으로 부여하기 위한 이미지 임베딩과 원본 데이터의 변형 조건이나 생성 대상 이미지의 컨텐츠 x를 텍스트 conditioning으로 주입하기 위한 텍스트 인코더로부터 출력되는 이미지 생성을 가이드 하기 위한 생성 조건으로 부여하기 위한 텍스트 임베딩이 연결(concatenate)되어 함께 주입되는 conditioning diffusion을 구현하는 diffusion 모델의 Denoising process로부터 가상 데이터($x^{\sim}$)로 구현된 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

도 8에는 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 화소 공간 보다 저차원의 레이턴트 공간(latent space) 상에서 디퓨전이 구현되는 레이턴트 디퓨전 모델(LDM, Latent Diffusion Model 또는 stable Diffusion Model)로 구현된 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 원본 데이터 및 가상 데이터(y)를 파인-튜닝의 학습 데이터로 하여, 원본 데이터 또는 가상 데이터(y)로부터 Noising process 및 Denoising process를 구현하면서 원본 데이터 또는 가상 데이터(y)로부터 복원 데이터($y^{\sim}$)를

생성하도록 파인-튜닝을 구현하는 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

도 9에는 각각 쌍을 이루는 다수의 이미지와 상기 다수의 이미지를 설명하는 captioning에 해당되는 텍스트를 입력으로 하여, 서로 매칭되는 텍스트 인코더와 이미지 인코더를 학습시키도록 각각의 텍스트 인코더로부터 출력된 텍스트 임베딩과 이미지 인코더로부터 출력된 이미지 임베딩 사이에서 서로 매칭되는 쌍 사이에서는 코사인 유사도를 최대로 하고, 서로 매칭되지 않는 쌍에서는 코사인 유사도를 최소로 하도록 학습시키는 contrastive pre-training(contrastive language image pre-training, CLIP)을 설명하기 위한 도면이 도시되어 있다.

도 10에는 본 발명의 일 실시형태에서, 레이턴트 표현(random vector, latent code)를 입력으로 하여 서로 다른 속성들 간의 disentanglement를 위한 non-linear mapping network(FC layer의 적층)와 mapping network로부터 산출된 intermediate 레이턴트 표현(w, intermediate latent representation) 또는 mapping network의 출력과 스타일 데이터(y)로부터 산출된 intermediate 레이턴트 표현(w, intermediate latent representation)과 컨텐츠 데이터(x)를 입력으로 하여, 스타일 데이터(y)로부터 전이된 스타일로 컨텐츠 데이터의 컨텐츠가 표현된 데이터를 생성하기 위한, StyleGAN을 설명하기 위한 도면이 도시되어 있다.

도 11에는 도 10에 도시된 StyleGAN의 네트워크에서 synthesis network g를 형성하는 것으로 컨텐츠 데이터(x)로부터 추출된 고정된 스케일의 Const 4x4x512 레이어와, 랜덤한 생성을 위한 stochastic variation 또는 stochastic detail을 생

성하기 위하여 각각의 스케일에 noise를 주입하기 위한 B layer(Per pixel noise injection)와, 스타일 데이터(y)의 feature statistics로부터 추출된 스케일링 파라메타 및 바이어스 파라메타를 적용하여 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature를 instant normalization하기 위한 AdaIN(Adaptive Instant Normalization)을 각각 설명하기 위한 도면이 도시되어 있다.

도 12에는 도 10에 도시된 StyleGAN 네트워크의 아키텍처를 보다 간략하게 개략적으로 표현한 것으로, 저해상도 이미지(또는 저해상도 feature map)로부터 고해상도 이미지를 향하여 progressive하게 업 스케일링으로 전개되는 멀티-스케일을 포함하는 synthesis network를 설명하기 위한 도면이 도시되어 있다.

도 13에는 도 10에 도시된 StyleGAN 네트워크의 아키텍처에서, 저해상도 이미지(또는 저해상도 feature map)로부터 고해상도 이미지를 향하여 progressive하게 업 스케일링으로 전개되는 멀티-스케일의 각각의 스케일에서 학습 내지는 파인-튜닝이 구현되는 StyleGAN 네트워크를 설명하기 위한 도면이 도시되어 있다.

도 14에는 본 발명의 일 실시형태에서, 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 모델을 설명하기 위한 도면으로, 레이턴트 표현(z)을 입력으로 하는 Mapping network의 출력과 스타일 데이터(y)를 입력으로 하여 intermediate 레이턴트 표현(w)을 추출하고 레이턴트 표현으로부터 추출된 스타일 정보(y^)을 각각의 멀티 스케일에 주입하면서, 컨텐츠 데이터(x)를 입력으로 하여 저해상도의 이미지(또는 저해상도 feature map)로부터 고해상도 이미지를 향하여 업 스케일을 구현하면서, 스타일 데이터(y)로부터 전이된 스타일로, 컨텐츠 데이터의 컨텐츠가 표

현된 가상 이미지를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

도 15에는 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 모델을 설명하기 위한 도면으로, 스타일 데이터(y)와 컨텐츠 데이터(x)를 입력으로 하여, 스타일 데이터(y)로부터 전이된 스타일로, 컨텐츠 데이터(x)의 컨텐츠가 표현된 합성 데이터가 생성되도록, 파인-튜닝을 구현하기 위한 제2 디지털 미디어 컨텐츠 모델을 설명하기 위한 도면으로, style loss(Ls)와 contents loss(Lc)를 산출하기 위한 loss computing 네트워크(f,ϕ)를 포함하는 제2 디지털 미디어 컨텐츠 모델이 도시되어 있다.

도 16에는 컨텐츠 데이터(x)와 스타일 데이터(y)로부터 이들 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature(또는 feature map)를 추출하고 차원을 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 스타일 데이터(y)로부터 추출된 feature statistic로부터 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여 instant normalization 시키기 위한 AdaIN(Adaptive Instant Normalization)과, 인코더로부터 차원 축소된 feature(또는 AdaIN이 적용된 feature)를 입력으로 하여 차원 확장시키면서 고해상도의 이미지를 생성하기 위한 디코더(Decoder)를 포함하며, 학습 또는 파인-튜닝을 위한 style loss(Ls) 및 contents loss(Lc)를 산출하기 위한 loss computing 네트워크(예를 들어, VGG Encoder)를 포함하는 인코더-디코더 아키텍처의 Style Transfer

네트워크를 설명하기 위한 도면이 도시되어 있다.

도 17에는 본 발명의 일 실시형태에서, 컨텐츠 데이터(x)와 스타일 데이터(y)를 입력으로 하여 가상 데이터를 생성하기 위한 제1 컨텐츠 미디어 생성 모델을 설명하기 위한 도면으로, 컨텐츠 데이터(x)와 스타일 데이터(y)로부터 이들 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature(또는 feature map)를 추출하고 차원을 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 스타일 데이터(y)로부터 추출된 feature statistic로부터 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여 instant normalization 시키기 위한 AdaIN(Adaptive Instant Normalization)과, 인코더로부터 차원 축소된 feature(또는 AdaIN이 적용된 feature)를 입력으로 하여 차원 확장시키면서 고해상도의 이미지를 생성하기 위한 디코더(Decoder)를 포함하는 Style Transfer 네트워크로 구현된 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

도 18에는 본 발명의 일 실시형태에서 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 컨텐츠 데이터(x)와 스타일 데이터(y)로부터 이들 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature(또는 feature map)를 추출하고 차원을 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 스타일 데이터(y)로부터 추출된 feature statistic로부터 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여

instant normalization 시키기 위한 AdaIN(Adaptive Instant Normalization)과, 인코더로부터 차원 축소된 feature(또는 AdaIN이 적용된 feature)를 입력으로 하여 차원 확장시키면서 고해상도의 이미지를 생성하기 위한 디코더(Decoder)를 포함하고, 학습 또는 파인-튜닝을 위한 style loss(Ls) 및 contents loss(Lc)를 산출하기 위한 loss computing 네트워크(f,ϕ)를 포함하는 Style Transfer 네트워크를 설명하기 위한 도면이 도시되어 있다.

도 19a 및 도 19b에는 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에서, 파라메타 행렬(W)에 대한 파인-튜닝을 통한 파라메타 행렬의 조정분(ΔW)에 대해 저차원 행렬 분해(low rank decomposition)로부터 저차원 행렬인 LoRA 행렬(LoRA matrices A,B 또는 LoRA 어뎁터 A,B)의 행렬 곱의 형태로 근사시키고, 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬(W)은 고정(freeze)시키는 LoRA(Low Rank Adaptation, 도 19a)와 제2 디지털 미디어 컨텐츠 모델의 파라메타를 업-데이트 시키는 통상적인 파인-튜닝(Weight update in regular fine-tuning, 도 19b)을 설명하기 위한 도면이 각각 도시되어 있다.

도 20에는 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬(W)에 대한 파인-튜닝을 통한 파라메타 행렬의 조정분(ΔW)에 대해 저차원 행렬 분해(low rank decomposition)로부터 저차원 행렬인 LoRA 행렬(LoRA matrices A,B 또는 LoRA 어뎁터 A,B)의 행렬 곱의 형태로 근사시키는 LoRA를 설명하기 위한 도면이 도시되어 있다.

도 21에는 도 19a에 도시된 LoRA를 보다 구체적으로 설명하기 위한 것으로,

1) 제2 디지털 미디어 컨텐츠 모델의 파라메타 고정(freeze), 2) 파인-튜닝을 위한 학습 데이터(원본 데이터+가상 데이터)가 제2 디지털 미디어 컨텐츠 모델과 LoRA adpater(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)에 함께 입력되는 Input, 3) adapter A(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)를 통과한 학습 데이터가 adapter B(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)를 통과하면서 BA의 행렬 곱이 생성되는 Adapter Train, 4) 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬 W를 통과한 출력과 LoRA adapter를 통과한 출력이 합산되면서 최종 출력의 산출을 설명하기 위한 도면이 도시되어 있다.

도 22에는 본 발명의 일 실시형태에서, 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에 적용되는 LoRA(Low Rank Adaptation)를 설명하기 위한 도면으로, 이전 시간 스텝에서의 은닉 상태(Hidden state)의 서로 다른 시간 스텝에서의 출력에 대해 쿼리 행렬(W_Q), 키 행렬(W_K) 및 밸류 행렬(W_V)을 승산하여 각각의 쿼리 벡터(Q), 키 벡터(K) 및 밸류 벡터(V)를 산출하는 프로세스에서 LoRA 어뎁터가 적용되는 것을 설명하기 위한 도면이 도시되어 있다.

도 23에는 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에 적용될 수 있는 QLoRA(Quantized Low Rank Adaptation)를 설명하기 위한 도면으로, 32비트 부동 소수점(float 32, FP32)의 제2 디지털 미디어 컨텐츠 모델(Before Quantization)의 파라메타를 4비트 NormalFloat(NF4)라는 새로운 데이터 타입으로 양자화시키는(After Quantization) QLoRA(Quantized Low Rank Adaptation)를 설명하기 위한 도면이 도시되어 있다.

도 24a 내지 도 24c에는 본 발명의 일 실시형태에 따른 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성의 일 예시로서, 서로 다른 원본 데이터(원본 데이터로서의 이미지)를 예시적으로 보여주는 도면들이 도시되어 있다.

도 25a 내지 도 25i에는 본 발명의 일 실시형태에 따른 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성의 일 예시로서, 도 24a에 예시된 원본 데이터(원본 데이터로서 이미지)에 부여된 서로 다른 변형 조건으로부터 생성된 서로 다른 가상 데이터(가상 데이터로서 이미지)를 예시적으로 보여주는 도면들이 도시되어 있으며, 보다 구체적으로, 도 25a 내지 도 25i에 예시된 각각의 서로 다른 가상 데이터(가상 데이터로서 이미지)는 서로 다른 변형 조건으로, 1) brightness_down, 2) brightness_up, 3) gray scale, 4) horizontal_flip, 5) rotation_+45 degree, 6) rotation_-45 degree, 7) saturation_down, 8) saturation_up, 9) vertical_flip의 변형 조건으로부터 생성된 서로 다른 가상 데이터를 보여주는 도면들이 도시되어 있다.

도 26a 내지 도 26c에는 각각 도 25a 내지 도 25c에 예시된 가상 데이터(가상 데이터로서 이미지)를 합성 데이터(합성 데이터로서 이미지)의 이미지 생성을 가이드 하기 위한 이미지 conditioning으로 하고, 합성 데이터(합성 데이터로서 이미지)의 이미지 생성을 가이드 하기 위한 텍스트 conditioning을 주입하여 생성된 합성 데이터를 예시적으로 보여주는 도면들로서, 하기와 같은 이미지 conditioning 및 텍스트 conditioning을 디퓨전 모델(Diffusion Model, 또는 레이턴트 디퓨전 모델, LDM, Latent Diffusion Model, 또는 스테이블 디퓨전 모델, stable Diffusion

Model, 예를 들어, 도 7에 도시된 제1 디지털 미디어 컨텐츠 생성 모델)에 주입하여 생성된 서로 다른 합성 데이터(합성 데이터로서 이미지)를 예시적으로 보여주는 도면이 도시되어 있다.

도 26a:

(이미지 conditioning) 도 24a의 가상 데이터

(텍스트 conditioning) "A mesmerizing abstract constellation floating in deep cosmic blues and purples, stars and swirling nebulae forming intricate patterns across a vast universe"

도 26b:

(이미지 conditioning) 도 24b의 가상 데이터

(텍스트 conditioning) "A vibrant abstract depiction of constellations interwoven with shimmering stardust, set against a backdrop of swirling galactic colors and cosmic haze"

도 26c:

(이미지 conditioning) 도 24c의 가상 데이터

(텍스트 conditioning) "An ethereal tapestry of stars and geometric constellation shapes, floating in a radiant universe painted with deep blues, violets, and bursts of light"

【발명을 실시하기 위한 구체적인 내용】

【0077】 이하, 첨부된 도면을 참조하여, 본 발명의 바람직한 실시형태에 관한 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성 시스템에 대해 설명하기로 한다.

【0078】 도 1에는 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 원본 데이터(스타일 데이터, 이미지 1)와 컨텐츠 데이터(이미지 2) 또는 원본 데이터의 변형 조건이나 컨텐츠 정보(텍스트)를 입력으로 하여, 가상 데이터(이미지 3, 전이된 스타일로 컨텐츠가 표현된 가상 이미지)를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델(생성형 AI 네트워크)을 설명하기 위한 도면이 도시되어 있다.

【0079】 도 2에는 본 발명의 일 실시형태에서, 원본 데이터(스타일 데이터, 이미지 4) 및 가상 데이터(스타일 데이터, 이미지 4)를 파인-튜닝의 학습 데이터로 하여, 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 모델(생성형 AI 모델)을 설명하기 위한 도면으로, 원본 데이터(스타일 데이터, 이미지 4) 또는 가상 데이터(스타일 데이터, 이미지 4)와 컨텐츠 데이터(이미지 5)를 입력으로 하여, 전이된 스타일로 컨텐츠가 표현된 합성 데이터를 생성하도록 파인-튜닝을 구현하는 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

【0080】 도 3a 및 도 3b에는 본 발명의 일 실시형태에서, 원본 데이터의 변형(transformation)을 통한 가상 데이터의 생성 또는 데이터 증강을 설명하기 위한 도면으로, 각각 (a) Flip, (b) Gray scale, (c) Saturation, (d) Brightness, (e) Rotation의 Img2Img transformation을 설명하기 위한 도면들이 도시되어 있다.

【0081】본 발명의 일 실시형태에 따른 데이터 증강(data augmentation)에 기반한 디지털 미디어 컨텐츠 생성 시스템은, 한정된 개수의 원본 데이터로부터 원본 데이터(ex. 디지털 미디어 컨텐츠로서 이미지)의 스타일을 추종하는 가상 데이터를 생성하여 데이터 증강을 구현하는 제1 디지털 미디어 컨텐츠 생성 모델과, 한정된 개수의 원본 데이터와 제1 디지털 미디어 컨텐츠 생성 모델로부터 생성된 가상 데이터를 취합한 학습 데이터로부터 학습(training) 또는 파인-튜닝(fine-tuning)된 파라메타를 포함하는 생성형 AI 모델에 대해 생성 조건(예를 들어, 합성 데이터의 생성을 가이드 하기 위한 텍스트의 생성 조건 또는 합성 데이터의 생성을 가이드 하기 위한 이미지의 생성 조건)을 주입하여 생성 조건을 추종하는 합성 데이터를 생성하기 위한 제2 디지털 미디어 컨텐츠 생성 모델을 포함할 수 있다.

【0082】예를 들어, 본 발명의 일 실시형태에서 상기 제1 디지털 미디어 컨텐츠 생성 모델을 통하여 생성된 가상 데이터는 원본 데이터의 희귀성의 한계에 따라 AI 기반으로 생성된 가상 데이터를 의미할 수 있으며, 예를 들어, 유사한 스타일의 디지털 미디어 컨텐츠에 대한 공급의 어려움 내지는 희귀성에도 불구하고 충분한 학습 데이터를 확보하고 확보된 학습 데이터로부터 과적합(over-fitting)이 없이 일반화 능력을 향상시키기 위하여 충분한 학습 데이터가 확보될 필요가 있으나, 예를 들어, 본 발명의 일 실시형태에서와 같이 스타일 전이(style transfer)를 위하여 동일 또는 유사한 스타일의 디지털 미디어 컨텐츠를 충분히 확보하는 것은 희귀성의 한계(취합 또는 데이터 수집에 어려움)가 있을 수 있으므로, 본 발명의 일 실시형태에서는 가용한 수준의 디지털 미디어 컨텐츠를 원본 데이터로 하여 원

본 데이터의 스타일 내지는 분포(distribution)를 추종하는 원본 데이터와 유사한 스타일 내지는 분포의 가상 데이터를 생성하는 데이터 증강(data augmentation)을 통하여 동일 또는 유사한 스타일 내지는 분포를 갖는 다수의 학습 데이터를 확보할 수 있다.

【0083】본 발명의 일 실시형태에서 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠의 생성에서는, 이하와 같은 서로 다른 데이터 증강 기법을 적용할 수 있다. 예를 들어, 본 발명의 일 실시형태에서 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠의 생성에서는 원본 데이터의 변형을 통한 데이터 증강이 적용될 수 있으며, 예시적으로, Brightness, Adjustment, Rotation, Flip, Crop, Affine transformation 등의 원본 데이터의 변형이 적용될 수 있으며, 보다 구체적으로, 원본 데이터를 대칭시키는 데이터 변형으로서, 예를 들어, 좌우 또는 상하로 원본 데이터를 반전시키는 Flipping, R,G,B 원본 데이터를 Gray scale로 변환해주는 Gray scale, R,G,B 원본 데이터를 HSV(Hue 색조, Saturation 채도, Value 명도)로 색감을 표현해주고 S 채널(Saturation 채널)에 오프셋(offset)을 적용하여 이미지의 선명도를 높이는 Saturation, 원본 데이터의 밝기를 조정해주는 Brightness, 원본 데이터의 각도를 변환해주는 Rotation, 원본 데이터의 사이즈를 변경해주는 Resize, 원본 데이터의 중앙을 기준으로 확대해주는 Center Crop, 원본 데이터에 대해 affine matrix를 적용하는 affine transformation 등의 원본 데이터의 변형이 적용될 수 있다.

【0084】<Diffusion - stable Diffusion>

【0085】도 4에는 디퓨전의 forward process로서, 특정한 패턴의 선명한 원본 이미지(swiss roll)에 시간의 스텝에 따라 점진적으로 노이즈 스케줄에 따른 노이즈를 추가하면서 원본 이미지의 특정한 패턴이 붕괴된 완전한 가우시안 노이즈(isotropic Gaussian noise)를 생성하는 noising process를 설명하기 위한 도면이 도시되어 있다.

【0086】도 5에는 디퓨전의 reverse process로서, 완전한 가우시안(isotropic Gaussian noise)로부터 예측된 노이즈를 제거하면서 시간의 스텝에 따라 덜 노이지한 이미지를 생성하면서, 원본 이미지의 패턴을 복원하는 이미지 생성을 설명하기 위한 도면으로, denoising process를 설명하기 위한 도면이 도시되어 있다.

【0087】도 6에는 노이즈가 추가된 이미지와, 해당되는 이미지의 시간 스텝을 입력으로 하여, 현재 시간 스텝에서 추가된 노이즈를 예측함으로써, 현재 시간 스텝에서의 노이지한 이미지로부터 상대적으로 덜 노이지한 이미지의 확률 분포를 예측하기 위한 U-net의 아키텍처를 설명하기 위한 도면이 도시되어 있다.

【0088】도 7에는 본 발명의 일 실시형태에서 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 화소 공간 보다 저차원의 레이턴트 공간(latent space) 상에서 디퓨전이 구현되는 레이턴트 디퓨전 모델(LDM, Latent Diffusion Model 또는 stable Diffusion Model)로 구현된 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 원본 데이터(y)를 이미지 conditioning으로 주입하기 위한 이미지 인코더로부터 출력되는 이미지 생성을 가이드 하기 위한

생성 조건으로 부여하기 위한 이미지 임베딩과 원본 데이터의 변형 조건이나 생성 대상 이미지의 컨텐츠 x를 텍스트 conditioning으로 주입하기 위한 텍스트 인코더로부터 출력되는 이미지 생성을 가이드 하기 위한 생성 조건으로 부여하기 위한 텍스트 임베딩이 연결(concatenate)되어 함께 주입되는 conditioning diffusion을 구현하는 diffusion 모델의 Denoising process로부터 가상 데이터($\tilde{x}$)로 구현된 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

【0089】 도 8에는 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 화소 공간 보다 저차원의 레이턴트 공간(latent space) 상에서 디퓨전이 구현되는 레이턴트 디퓨전 모델(LDM, Latent Diffusion Model 또는 stable Diffusion Model)로 구현된 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 원본 데이터 및 가상 데이터(y)를 파인-튜닝의 학습 데이터로 하여, 원본 데이터 또는 가상 데이터(y)로부터 Noising process 및 Denoising process를 구현하면서 원본 데이터 또는 가상 데이터(y)로부터 복원 데이터($\tilde{y}$)를 생성하도록 파인-튜닝을 구현하는 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

【0090】 도 9에는 각각 쌍을 이루는 다수의 이미지와 상기 다수의 이미지를 설명하는 captioning에 해당되는 텍스트를 입력으로 하여, 서로 매칭되는 텍스트 인코더와 이미지 인코더를 학습시키도록 각각의 텍스트 인코더로부터 출력된 텍스트 임베딩과 이미지 인코더로부터 출력된 이미지 임베딩 사이에서 서로 매칭되는 쌍 사이에서는 코사인 유사도를 최대로 하고, 서로 매칭되지 않는 쌍에서는 코사인

유사도를 최소로 하도록 학습시키는 contrastive pre-training(contrastive language image pre-training, CLIP)을 설명하기 위한 도면이 도시되어 있다.

【0091】본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠의 생성에서는 원본 데이터의 분포 내지는 확률 분포를 학습한 생성형 AI 모델(generative model)로부터 가상 데이터를 생성하는 데이터 증강이 적용될 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 제1 디지털 미디어 컨텐츠 생성 네트워크는 디퓨전(diffusion) 모델을 포함할 수 있다.

【0092】도 4 내지 도 8을 참조하면, 원본 데이터로부터 가상 데이터를 생성하기 위한 디퓨전에서는 특정한 패턴을 갖는 원본 이미지로부터 이산적인 시간 스텝을 전진시키면서 이전 시간 스텝의 이미지에 대해 각 시간 스텝에 대해 노이즈 스케줄로부터 정의된 가우시안 노이즈($\varepsilon \sim N(0,1)$, random noise)를 추가하면서 원본 이미지의 특정한 패턴이 소멸되도록 가우시안 확률 분포의 마르코프 체인(iterative Markov chain)을 포함하는 forward process로서 diffusion process와, 상기 forward process에 대한 reverse process로서 이전 시간 스텝으로부터 추가된 가우시안 노이즈($\varepsilon \sim N(0,1)$, random noise)를 제거하면서 원본 이미지의 특정한 패턴이 복원되도록 가우시안 확률 분포의 마르코프 체인(iterative Markov chain)을 포함하는 reverse process로서 denoising process를 정의할 수 있으며, forward process에서 노이즈가 추가된 이미지(Xt)와 해당되는 이미지의 시간 스텝(t=0,1,..,T)을 입력으로 하여, 추가된 노이즈를 예측하도록 학습될 수 있으며, 추

가된 노이즈를 제거해나가는 reverse process로부터 원본 이미지의 특정한 패턴을 복원시키는 이미지 생성(sampling process)을 수행할 수 있다.

【0093】보다 구체적으로, 상기 forward process(diffusion process)로서, 입력된 이미지(X0, 특정한 패턴이 선명한 원본 이미지)에 대해 시간 스텝을 전진(t=0,1,..,T)시키면서 순차적으로 정해진 노이즈 스케줄(βt, 예를 들어, 선형 스케일에 따라 점진적으로 증가되는 노이즈, fixed noise schedule)에 따라 가우시안 노이즈(Gaussian noise)를 추가하면서 사전에 설정된 시간 스텝(예를 들어, 1000개의 시간 스텝, T=1000)만큼 forward process를 수행하면, 사전에 설정된 시간 스텝의 이미지(XT)는 모든 방향에서 isotropic Gaussian으로 수렴하면서 원본 이미지의 특정한 패턴이 소멸된 완전한 가우시안 노이즈(normal Gaussian distribution, 평균 0, 분산 1, N(0,1))를 생성할 수 있다. 예를 들어, 상기 forward process는 이전 시간 스텝에서의 이미지(Xt-1)를 조건부로 하는 현재 시간 스텝의 이미지(Xt)에 대한 조건부 확률 분포의 체인(joint distribution)으로 표현될 수 있으며, 현재 시간 스텝(Xt)에서의 상태가 이전의 시간 스텝(Xt-1)에서의 상태에만 의존하는 마르코프 프로세스(Markov process)의 체인(joint distribution)으로 표현될 수 있다. 보다 구체적으로, 상기 forward process에서, 이전 시간 스텝에서의 이미지(Xt-1)를 조건부로 하는 현재 시간 스텝의 이미지(Xt)에 대한 조건부 확률 분포 q(Xt | Xt-1)는 이하와 같은 평균(mean) 및 분산(variance)을 갖는 조건부 가우시안 확률 분포로 정의될 수 있다.

【0094】

【0095】상기 forward process에서는 노이즈 스케줄(βt)에 따라 이전 시간 스텝에서의 값(Xt-1, 예를 들어, 화소 값 등)을 감소시키면서 가우시안 노이즈(ε~N(0,1), random noise)를 추가하여 현재 시간 스텝(Xt)에서의 이미지를 생성할 수 있다.

【0096】

【0097】상기 reverse process(denoising process)에서는 완전한 가우시안 노이즈(XT)로부터 원본 이미지(X0)를 복원하는 것으로, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델에서의 학습 대상에 해당될 수 있으며, 가상 데이터로서 이미지 생성을 위한 샘플링에 해당될 수 있다. 보다 구체적으로, 상기 reverse process(denoising process, sampling process)에서는 완전한 가우시안 노이즈(XT)로부터 점차적으로 가우시안 노이즈를 제거하면서 원본 이미지(X0)의 특정한 패턴을 복원할 수 있으며, 이러한 reverse process는 현재 시간 스텝에서의 이미지(Xt)를 조건부로 하는 이전 시간 스텝의 이미지(Xt-1)에 대한 조건부 확률 분포의 체인(joint distribution)으로 표현될 수 있으며, 현재 시간 스텝(Xt)에서의 상태가 이전의 시간 스텝(Xt-1)에서의 상태에만 의존하는 마르코프 프로세스(Markov process)의 체인(joint distribution)으로 표현될 수 있다. 상기 reverse process에서, 이전 시간 스텝에서의 이미지(Xt-1)를 조건부로

하는 현재 시간 스텝의 이미지(Xt)에 대한 조건부 확률 분포 q(Xt-1|Xt)는 노이즈 스케줄(βt)에 따라 노이즈가 추가되는 forward process와 달리, forward process의 조건부 확률 분포 q(Xt|Xt-1)로부터 조건부 시점이 역전된 조건부 확률 분포를 직접적으로 산출하기는 어렵기 때문에, reverse process의 조건부 확률 분포 q(Xt-1|Xt)는 파라메타(Θ)를 갖는 신경망(network) 모델을 통하여 이하와 같은 reverse process의 조건부 확률 분포 P_θ(Xt-1|Xt)로 근사(approximation)될 수 있다.

【0098】

【0099】상기 reverse process는 추론(inference) 단계에서 이미지 생성(sampling process)에 해당될 수 있고, 상기 reverse process는 이미지 생성을 위한 신경망 모델의 학습 대상에 해당될 수 있으며, 상기 이미지 생성을 위한 신경망 모델은 reverse process 조건부 확률 분포 q(Xt-1|Xt)를 예측하도록 학습될 수 있고, reverse process의 조건부 확률 분포 q(Xt-1|Xt) 중에서 분산은 고정하고(예를 들어, $\sigma_t^2=\beta t$로 상정) 평균(μ_θ(Xt,t))을 예측하도록 학습될 수 있으며, 예를 들어, reverse process의 조건부 확률 분포 q(Xt-1|Xt)와 신경망 모델을 통하여 근사되는 reverse process의 조건부 확률 분포 P_θ(Xt-1|Xt) 사이에서 KL divergence가 최소가 되도록 신경망 모델의 학습을 위한 목적 함수(objective function)가 유도될 수 있으며, 신경망 모델로부터 예측된 평균과 forward process

에서 산출된 실제 평균(예를 들어, 타겟 레이블) 사이의 오차를 최소화시키도록, 예를 들어, 이들 오차를 목적 함수(objective function) 내지는 손실 함수(loss function)로 하여, 목적 함수 내지는 손실 함수의 디센트 그라디언트(descent gradient)로부터 reverse process를 근사하는 신경망 모델의 파라메타를 갱신할 수 있다. 하기에서 E는 기대 값(expectation)을 의미할 수 있다.

【0100】

【0101】 이때, 평균은 이하와 같이 표현될 수 있으며, reverse process를 근사하는 신경망 모델은 현재의 시간 스텝에서의 이미지(Xt)로부터 덜 노이지(noisy)한 이전 시간 스텝에서의 이미지(Xt-1)의 평균을 직접 예측하도록 학습된다기 보다는, 이전 시간 스텝에서의 이미지(Xt-1)로부터 추가된 가우시안 노이즈 ε_θ(Xt,t)를 예측함으로써, 현재의 시간 스텝에서의 덜 노이지한 이미지(Xt)의 평균을 예측할 수 있으며, 상기와 같이, 신경망 모델로부터 예측된 평균과 forward process에서 산출된 실제 평균(타겟 레이블)과의 오차를 최소화시키도록 신경망 모델의 목적 함수가 구성될 수 있다.

【0102】

【0103】

【0104】 여기서,

【0105】

【0106】

【0107】 여기서,

【0108】

【0109】 상기와 같은 reverse process를 근사하는 신경망 모델의 학습에서는 이하와 같은 그래디언트 디센트를 취하여 신경망 모델의 파라메타를 갱신할 수 있다.

【0110】

【0111】 그리고, 상기와 같이 학습된 신경망 모델로부터 예측된 가우시안 노이즈로부터 노이즈가 더한 현재 시간 스텝에서의 이미지(Xt)로부터 노이즈가 덜한 이전 시간 스텝에서의 이미지(Xt-1)를 생성할 수 있으며, 보다 구체적으로, 현재

시간 스텝에서의 이미지(Xt)로부터 예측된 가우시안 노이즈($\varepsilon_\theta(xt,t)$)로부터 노이즈가 덜한 이전 시간 스텝에서의 이미지(Xt-1)의 평균을 산출할 수 있고, 고정된 분산($\sigma_t^2=\beta t$, 예를 들어, 노이즈 스케줄 βt과 같은 값으로 상정) 및 샘플링을 위한 가우시안 노이즈(Z~N(0.1))로부터 이하와 같이 산출될 수 있다.

【0112】

【0113】 여기서,

【0114】

【0115】 상기 reverse process를 근사하기 위한 신경망 모델은, 상대적으로 노이지가 더한 현재 시간 스텝에서의 이미지(Xt)로부터 상대적으로 노이지가 덜한 이전 시간 스텝에서의 이미지(Xt-1)의 평균 및 분산을 예측함으로써(현재 시간 스텝에서의 이미지 Xt로부터 추가된 가우시안 노이즈 $\varepsilon_\theta(xt,t)$의 예측을 통하여 이전 시간 스텝에서의 이미지 Xt-1의 평균을 예측함), 현재 시간 스텝에서의 이미지를 조건부로 하여 이전 시간 스텝에서의 이미지의 확률 분포를 예측할 수 있으며, 보다 원본 이미지에 가까운, 이전 시간 스텝에서의 이미지를 생성하여 원본 이미지의 특정한 패턴을 복원할 수 있으며, 이미지 생성을 위한 모델로 적용될 수 있다.

【0116】예를 들어, 현재 시간 스텝의 이미지와 현재의 시간 스텝을 입력으로 하여, 상대적으로 덜 노이지한 이전 시간 스텝에서의 이미지에 추가된 가우시안 노이즈를 예측하기 위한 신경망 모델, 달리 표현하면, 상대적으로 더 노이지한 현재 시간 스텝에서의 이미지를 조건부로 하여, 상대적으로 덜 노이지한 이전 시간 스텝에서의 이미지의 확률 분포를 예측하기 위한 신경망 모델은 U-net 구조의 아키텍처로 형성될 수 있다. 상기 U-net 구조의 아키텍처는 입력된 고차원의 이미지(예를 들어, 노이지한 영상)로부터 영상의 특성 맵 또는 특징을 추출하도록 다운 샘플링을 통하여 저차원의 영상을 생성하기 위한 수축 경로(contracting path)와 저차원의 영상으로부터 행렬 차원을 증가시키는 업 샘플링을 통하여 입력된 이미지와 같은 고차원의 영상을 생성하기 위한 확장 경로(expanding path)를 포함할 수 있다. 본 발명의 일 실시형태에서, 조건부 확률 분포 $q(Xt-1 | Xt)$를 근사하는 조건부 확률 분포 $P_\theta(Xt-1 | Xt)$를 예측하기 위한 신경망 모델은, 현재 시간 스텝에서의 이미지와 함께 현재 시간 스텝을 입력으로 하여, 상대적으로 덜 노이즈한 이전 시간 스텝에서의 이미지의 조건부 확률 분포를 예측할 수 있으며, 이를 위해, 상기 현재의 시간 스텝(t)은 임베딩을 통하여 상기 U-net 아키텍처로 구현된 신경망 모델로 입력될 수 있으며, 예를 들어, U-net 아키텍처의 다수의 레이어로 입력될 수 있다. 후술하는 바와 같이, 조건부 확률 분포 $q(Xt-1 | Xt)$를 근사하는 조건부 확률 분포 $P_\theta(Xt-1 | Xt)$를 산출하기 위한 신경망 모델은, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델로서, 예를 들어, 가상 데이터로서 이미지 생성의 condition을 부여하는 텍스트와 같은 conditioning이 주입

될 수 있으며, 본 발명의 일 실시형태에서 텍스트 conditioning이 텍스트 인코더를 통하여 텍스트 임베딩으로 입력될 수 있고, 입력된 텍스트 임베딩과의 관련성을 높이도록, 생성 대상 이미지의 데이터(레이턴트 표현)를 쿼리(query)로 하고 입력된 텍스트 임베딩을 키(key)와 밸류(value)로 하는 크로스 어텐션(cross-attention)을 수행하기 위한 어텐션 구조를 포함할 수 있다.

【0117】상기 디퓨전이 적용된 제1 디지털 미디어 컨텐츠 생성 모델은, 디지털 미디어 컨텐츠로서 이미지의 화소 공간(pixel space) 상에서 구현될 수도 있으나 화소 공간 보다 저차원의 레이턴트 공간(latent space) 상에서 구현될 수도 있으며(LDM, latent diffusion model, 또는 stable diffusion), 예를 들어, 다양성과 높은 질의 이미지 생성이 가능한 반면에, 다른 이미지 생성 모델(GAN, VAE, flow) 보다 상대적으로 이미지 생성에 소요되는 시간이 길다는 점을 감안하고, 또한, 디퓨전의 학습은 고주파 성분(high frequency detail)은 사라지지만 semantic은 유지되는 perceptual compression과 실제 데이터의 본질이 학습되는 semantic compression의 두 단계로 학습이 진행될 때, perceptual compression에서 상대적으로 긴 시간이 소요된다는 점에서, 디지털 미디어 컨텐츠로서 이미지의 화소 단위로 디퓨전(픽셀 스페이스 모델, pixel space model)을 수행하기 보다는, 저차원의 레이턴트 공간(latent space) 상에서 디퓨전을 수행할 수도 있으며(레이턴트 스페이스 모델, latent space model), 앞서 설명된 바와 같은 forward process와 reverse process가 저차원의 레이턴트 공간(latent space) 상에서 구현될 수 있다.

【0118】LDM(latent diffusion model) 또는 스테이블 디퓨전(stable diffusion)에서는 디퓨전의 forward process와 reverse process가 구현되는 신경망 블록의 전단 및 후단으로, 입력된 디지털 미디어 컨텐츠로서 이미지의 화소 공간을 레이턴트 공간으로 인코딩하기 위한 인코더(입력 이미지의 화소 데이터를 레이턴트 표현-latent representation-로 인코딩)와, 생성된 디지털 미디어 컨텐츠로서 이미지를 레이턴트 공간 상에서 화소 공간으로 디코딩하기 위한 디코더(생성된 이미지에 관한 레이턴트 표현-latent representation-을 화소 공간의 표현으로 디코딩)를 더 포함할 수 있으며, 상기 인코더와 디코더는 오토인코더(autoencoder)의 인코더 및 디코더의 쌍으로 구현될 수 있다. 상기 인코더와 디코더 사이의 레이턴트 공간 상에서는 forward process와 reverse process가 구현될 수 있고, 예를 들어, 앞서 설명된 바와 같이, 조건부 확률 분포 $q(Xt-1|Xt)$를 근사하는 조건부 확률 분포 $P_\theta(Xt-1|Xt)$를 산출하기 위한 U-net 아키텍처의 신경망 모델은 상기 인코더와 디코더 사이에 구현될 수 있다. 도 7에 도시된 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델에서, 인코더에 대한 입력과 디코더로부터의 출력은 화소 공간 상에서 이루어질 수 있으며, 인코더 및 디코더 사이에서 디퓨전의 forward process와 reverse process는 레이턴트 공간 상에서 구현될 수 있고, reverse process(denoising process) 상에서 현재 시간 스텝에서의 이미지와 현재의 시간 스텝을 입력으로 하여, 보다 덜 노이지한 이전 시간 스텝에서의 이미지(디지털 미디어 컨텐츠)를 생성하도록 조건부 확률 분포를 예측하기 위한(예를 들어, 가우시안 노이즈 $\varepsilon_\theta(Xt,t)$를 예측하여, 보다 덜 노이지한 이전 시간 스

텝에서의 평균을 예측함) U-net 아키텍처의 신경망 모델이 구현될 수 있다. 이때, 상기 reserve process(denoising process) 상의 U-net 구조에는 생성 이미지(생성 디지털 미디어 컨텐츠)에 관한 conditioning의 주입을 위한 텍스트 인코더가 연결될 수 있으며(conditioning diffusion), 본 발명의 일 실시형태에서, 상기 텍스트 인코더는 이미지-텍스트의 멀티 모달(multi-modal)의 임베딩 공간을 학습한 멀티 모달 AI 모델로 구현될 수 있으며, 예를 들어, 본 발명의 일 실시형태에서 멀티 모달 AI 모델로서, 클립 텍스트 임베딩(CLIP text embedding)을 위한 텍스트 인코더일 수 있다.

【0119】 예를 들어, 상기 텍스트 인코더는 입력된 텍스트로부터 이미지 임베딩과 매칭될 수 있는 텍스트 임베딩을 생성할 수 있으며, 텍스트 인코더로부터의 텍스트 임베딩은 생성 대상 디지털 미디어 컨텐츠로서 이미지와의 크로스 어텐션(cross-attention)을 통하여 가상 데이터로서 디지털 미디어 컨텐츠의 생성 또는 이미지 생성에 대한 conditioning을 주입할 수 있으며, 보다 구체적으로, 생성 대상 디지털 미디어 컨텐츠로서 이미지의 데이터(레이턴트 표현)를 쿼리(query)로 하고, 입력된 텍스트 임베딩을 키(key)와 밸류(value)로 하는 크로스 어텐션을 수행할 수 있다. 그리고, 상기 디퓨전의 reverse process와 디코더 사이에는 디퓨전의 reverse process로부터 예측된 가우시안 노이즈 $\varepsilon_\theta(Xt,t)$로부터 상대적으로 덜 노이지한 이전 시간 스텝에서의 이미지에 관한 조건부 확률 분포 $P_\theta(Xt-1 \mid Xt)$의 평균 및 분산으로부터 샘플링을 통하여 이전 시간 스텝에서의 덜 노이지한 이미지를 생성하기 위한 샘플링 프로세스(sampling process)가 연결될 수 있다.

【0120】상기 LDM(latent diffusion model) 또는 스테이블 디퓨전(stable diffusion)에서는, 학습을 위한 구성으로 추론(inference) 과정에서는 필요하지 않을 수 있는, 레이턴트 표현(latent representation)을 위한 인코더와 forward process(diffusion process)는 학습 이후의 신경망 모델의 경량화를 위하여 구비되지 않은 상태로 배포(deploy)될 수 있으며, 학습된 파라메타를 포함하는 reverse process로부터 현재 시간 스텝에서의 이미지 보다 덜 노이지한 이전 시간 스텝에서의 이미지의 확률 분포를 예측할 수 있다.

【0121】본 발명의 일 실시형태에 따른 제1 디지털 미디어 컨텐츠 생성 모델은, 상대적으로 많은 대규모 학습 데이터로부터 학습이 이루어진 training된 모델 또는 pre-training된 모델에 대해, 원본 데이터를 fine-tuning의 학습 데이터로 하여 fine-tuning된 모델로 구현될 수 있으며, 원본 데이터로부터 fine-tuning된 모델로서 제1 디지털 미디어 컨텐츠 생성 모델로부터 가상 데이터로서의 디지털 미디어 컨텐츠 또는 이미지의 생성 조건으로 원본 데이터에 대한 변형 조건을 가상 데이터의 생성 조건으로 주입할 수 있으며, 예를 들어, 원본 데이터의 종횡비 또는 사이즈를 변경해주는 Resize 또는 Crop and Resize 또는 설정된 해상도에 따라 원본 데이터의 종횡비 또는 사이즈를 변경해주면서 공백을 보충해주는(예를 들어, 최다 빈도의 화소 기반으로 바탕 처리) Resize and Fill, 원본 데이터의 각도를 변환해주는 Rotation, 원본 데이터의 마스킹 처리된 영역에 대한 object를 생성해주는 Inpaint 등과 같은 원본 데이터의 변경 조건을 가상 데이터의 생성 조건으로 하여 가상 데이터로서 디지털 미디어 컨텐츠 또는 이미지의 생성 조건으로 주입할 수 있

으며, 텍스트 인코더를 통하여 텍스트 임베딩의 형태로 conditioning이 주입될 수 있는 텍스트 형태로 입력될 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델은 가상 데이터의 생성 조건으로서 텍스트 형태의 conditioning의 주입을 위한 텍스트 인코더를 포함할 수 있으며(도 7 참조), 본 발명의 일 실시형태에서, 상기 텍스트 인코더는 이미지-텍스트의 멀티 모달(multi-modal)의 임베딩 공간을 학습한 멀티 모달 AI 모델로 구현될 수 있으며, 예를 들어, 본 발명의 일 실시형태에서 멀티 모달 AI 모델로서, 클립 텍스트 임베딩(CLIP text embedding)을 위한 텍스트 인코더를 포함할 수 있다(도 9 CLIP 참조).

【0122】<StyleGAN>

【0123】도 10에는 본 발명의 일 실시형태에서, 레이턴트 표현(random vector, latent code)를 입력으로 하여 서로 다른 속성들 간의 disentanglement를 위한 non-linear mapping network(FC layer의 적층)와 mapping network로부터 산출된 intermediate 레이턴트 표현(w, intermediate latent representation) 또는 mapping network의 출력과 스타일 데이터(y)로부터 산출된 intermediate 레이턴트 표현(w, intermediate latent representation)과 컨텐츠 데이터(x)를 입력으로 하여, 스타일 데이터(y)로부터 전이된 스타일로 컨텐츠 데이터의 컨텐츠가 표현된 데이터를 생성하기 위한, StyleGAN을 설명하기 위한 도면이 도시되어 있다.

【0124】도 11에는 도 10에 도시된 StyleGAN의 네트워크에서 synthesis network g를 형성하는 것으로 컨텐츠 데이터(x)로부터 추출된 고정된 스케일의 Const 4x4x512 레이어와, 랜덤한 생성을 위한 stochastic variation 또는

stochastic detail을 생성하기 위하여 각각의 스케일에 noise를 주입하기 위한 B layer(Per pixel noise injection)와, 스타일 데이터(y)의 feature statistics로부터 추출된 스케일링 파라메타 및 바이어스 파라메타를 적용하여 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature를 instant normalization하기 위한 AdaIN(Adaptive Instant Normalization)을 각각 설명하기 위한 도면이 도시되어 있다.

【0125】 도 12에는 도 10에 도시된 StyleGAN 네트워크의 아키텍처를 보다 간략하게 개략적으로 표현한 것으로, 저해상도 이미지(또는 저해상도 feature map)로부터 고해상도 이미지를 향하여 progressive하게 업 스케일링으로 전개되는 멀티-스케일을 포함하는 synthesis network를 설명하기 위한 도면이 도시되어 있다.

【0126】 도 13에는 도 10에 도시된 StyleGAN 네트워크의 아키텍처에서, 저해상도 이미지(또는 저해상도 feature map)로부터 고해상도 이미지를 향하여 progressive하게 업 스케일링으로 전개되는 멀티-스케일의 각각의 스케일에서 학습 내지는 파인-튜닝이 구현되는 StyleGAN 네트워크를 설명하기 위한 도면이 도시되어 있다.

【0127】 도 14에는 본 발명의 일 실시형태에서, 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 모델을 설명하기 위한 도면으로, 레이턴트 표현(z)을 입력으로 하는 Mapping network의 출력과 스타일 데이터(y)를 입력으로 하여 intermediate 레이턴트 표현(w)을 추출하고 레이턴트 표현으로부터 추출된 스타일 정보(y^)을 각각의 멀티 스케일에 주입하면서, 컨텐츠 데이터(x)를 입력으로 하여

저해상도의 이미지(또는 저해상도 feature map)로부터 고해상도 이미지를 향하여 업 스케일을 구현하면서, 스타일 데이터(y)로부터 전이된 스타일로, 컨텐츠 데이터의 컨텐츠가 표현된 가상 이미지를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

【0128】도 15에는 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 모델을 설명하기 위한 도면으로, 스타일 데이터(y)와 컨텐츠 데이터(x)를 입력으로 하여, 스타일 데이터(y)로부터 전이된 스타일로, 컨텐츠 데이터(x)의 컨텐츠가 표현된 합성 데이터가 생성되도록, 파인-튜닝을 구현하기 위한 제2 디지털 미디어 컨텐츠 모델을 설명하기 위한 도면으로, style loss(Ls)와 contents loss(Lc)를 산출하기 위한 loss computing 네트워크(f,ϕ)를 포함하는 제2 디지털 미디어 컨텐츠 모델이 도시되어 있다.

【0129】본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 생성 모델은, 생성형 AI 모델(generative model)로서, 앞서 설명된 바와 같은 디퓨전 모델(예를 들어, 레이턴트 디퓨전 모델 Latent Diffusion Model 또는 스테이블 디퓨전 모델 stable Diffusion Model)에 기반할 수 있으며, 예를 들어, 디퓨전 모델에 기반한 레이턴트 디퓨전 모델(LDM, Latent Diffusion Model) 내지는 스테이블 디퓨전 모델(stable diffusion model)을 포함할 수 있으며, 본 발명의 다양한 실시형태에서, 원본 데이터로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 모델은 또 다른 생성형 AI 모델(generative model)로서, GAN(Generative Adversarial Network) 기반의

StyleGAN(Style-Based Generator Architecture for Generative Adversarial Networks) 모델을 포함할 수 있다.

【0130】 상기 StyleGAN에서는 입력된 저차원의 레이턴트 표현(z, latent representation, 레이턴트 벡터, random vector, latent code, 1x512) 또는 입력된 저차원의 레이턴트 표현(z)을 입력으로 하여 mapping network로부터 추출된 intermediate 레이턴트 표현(w, intermediate latent representation, 예를 들어, 1x512)으로부터 convolution layer(Conv 3x3, 또는 convolution Neural Network, CNN 네트워크)를 통하여 출력된 출력 값에 대해 AdaIN(Adaptive instant Normalization)을 적용할 수 있다. 보다 구체적으로, 상기 StyleGAN에서는 컨텐츠 데이터(x), 예를 들어, 1024x1024의 컨텐츠 데이터(x)로부터 추출된 고정된 스케일(예를 들어, constant tensor 4x4x512로 추출된 feature map)의 컨텐츠 데이터(x)에 대해 convolution layer(Conv 3x3)를 적용할 수 있으며, convolution layer(Conv 3x3)을 출력에 대해 AdaIN(Adaptive instant Normalization)을 적용할 수 있다. 상기 AdaIN(Adaptive instant Normalization)에 관하여, 레이턴트 표현(레이턴트 벡터, random vector, latent code, 1x512)을 입력으로 하여 각각의 속성들 간의 disentanglement를 위한 non-linear mapping network의 출력과 스타일 데이터(y)를 입력으로 하여 intermediate 레이턴트 표현(w)를 추출할 수 있으며, 이와 같이 mapping network의 출력과 스타일 데이터(y)를 입력으로 하여 산출된 intermediate 레이턴트 표현(w)에 대한 learned affine transform(A)을 통하여 산출된 스타일 정보(y^)를 업-스케일링을 통하여 멀티-스케일로 progressive 하게 구

현되는 각각의 스케일에서 전이시키도록, 각각의 스케일에서 AdaIN(Adaptive instant Normalization)을 적용하여, 스타일 데이터(y)로부터 추출된 스타일(y^)이 전이되도록 할 수 있다.

【0131】보다 구체적으로, 본 발명의 일 실시형태에서 제1 디지털 미디어 컨텐츠 생성 모델로서 StyleGAN은, 레이턴트 표현(random vector, latent code, 1x512)을 입력으로 하여 서로 다른 속성들(서로 다른 특성들) 간의 disentanglement를 위한 non-linear mapping network을 포함할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서 상기 non-linear mapping network는 서로 적층된 8개의 FC layer(Fully Connected layer)의 연쇄를 포함할 수 있으며, 상기 mapping network로부터 산출된 intermediate 레이턴트 표현(w, intermediate latent representation, 예를 들어, 1x512) 또는 상기 mapping network의 출력과 스타일 데이터(y)를 입력으로 하여 산출된 intermediate 레이턴트 표현(w, intermediate latent representation, 예를 들어, 1x512)을 입력으로 하여 가상 데이터를 생성하기 위한 synthesis network를 포함할 수 있으며, 보다 구체적으로, 상기 synthesis network에서는 컨텐츠 데이터(x, 예를 들어, 1024x1012의 컨텐츠 데이터(x)로부터 고정된 스케일로 추출된 4x4x512의 constant tensor)를 입력으로 하고 저해상도로부터 고해상도로 progressive하게 업-스케일을 진행하면서 멀티-스케일을 형성하는 각각의 스케일에서 선행하는 프로세스를 통하여 스타일 데이터(y)로부터 추출된 스타일 정보(y^)를 AdaIN(Adaptive instant Normalization)를 통하여 전이시키면서, 예를 들어, coarse scale의 coarse style로부터 middle scale의 middle style, 그

리고 fine scale의 fine style에 이르기까지 서로 다른 스케일에서 서로 다른 스타일 정보(y^)를 전이시키면서, 저해상도의 이미지로부터 고해상도의 가상 이미지를 생성할 수 있다.

【0132】본 발명의 일 실시형태에서, 상기 StyleGAN은 레이턴트 표현(z, 레이턴트 벡터, random vector, latent code)을 입력으로 하여 intermediate 레이턴트 표현을 출력하기 위한 mapping network를 포함할 수 있으며, 예를 들어, 상기 mapping network는 8개의 서로에 대해 적층된 FC(fully connected layer)의 연쇄를 포함할 수 있다. 예를 들어, 상기 mapping network는 서로 다른 속성들(특징들, feature) 간의 entanglement를 방지하고 서로 다른 서로 다른 속성들(특징들, feature) 간의 disentanglement를 구현할 수 있으며, 예를 들어, Non-linear mapping network에 해당될 수 있다.

【0133】본 발명의 일 실시형태에서 제1 디지털 미디어 컨텐츠 생성 모델은, 레이턴트 표현(z, random vector, latent code) 또는 레이턴트 표현(z)을 입력으로 하여 feature 간의 disentanglement를 위한 mapping network를 통하여 생성된 intermediate 레이턴트 표현(w, intermediate latent representation)을 입력으로 하여 이미지 생성을 위한 synthesis network에서는, 컨텐츠 데이터(x) 내지는 고정된 스케일의 컨텐츠 데이터(x, constant tensor 4x4x512, 예를 들어, 1024x1024의 컨텐츠 데이터(x)로부터 추출된 고정된 스케일의 feature map)로부터 점진적으로 해상도를 증가시키면서 최종적으로 고해상도(예를 들어, 1024x1024의 고해상도)의 이미지(가상 데이터, 예를 들어, 디지털 미디어 컨텐츠로서의 이미지)를 생성할 수

있으며, 이와 같이 저해상도의 이미지로부터 고해상도의 이미지로 progressive 하게 업-스케일이 진행되는 다단의 멀티-스케일을 형성하는 각각의 스케일에서 스타일 컨텐츠(y, 예를 들어, 원본 데이터)로부터 추출된 서로 다른 스케일에 적용될 수 있는 스타일 정보(y^), 예를 들어, coarse scale의 coarse style로부터 middle scale의 middle style, fine scale의 fine style에 이르는 각각의 서로 다른 스케일에 적용될 수 있는 스타일 정보(y^)를 서로 다른 스케일에서 synthesis network를 통하여 생성되는 데이터로 전이(style transfer)시킬 수 있으며, 예를 들어, 스타일의 전이를 위한 AdaIN(Adaptive instant Normalization)을 통하여 스타일 컨텐츠(y, 예를 들어, 원본 데이터)로부터 추출된 서로 다른 스케일에서의 스타일 정보(y^)를 서로 다른 스케일에서 생성되는 데이터로 전이(style transfer)시킬 수 있으며, 상기 AdaIN(Adaptive instant Normalization)에서는 이하와 같은 AdaIN operation을 통하여 Instance Normalization을 구현할 수 있다.

【0134】

【0135】여기서, 각각의 x와 y는 컨텐츠 데이터(x)와 스타일 데이터(y)를 의미할 수 있으며, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 스타일 정보(y^) 또는 스타일 데이터(y)로부터 추출된 feature map의 mean(μ)과 standard variation(σ)을 스케일링(scaling) 파라메타 및 바이어스(bais) 파라메타로 적용하여 synthesis network에서 생성되는 데이터에 대해 instant normalization을 구

현할 수 있다.

【0136】본 발명의 일 실시형태에서, 상기 제1 디지털 미디어 컨텐츠 생성 모델로서 StyleGAN에서는 멀티-스케일을 형성하는 각각의 스케일에서 noise(B)를 주입(Per pixel noise injection)해줌으로써, 랜덤한 생성이 필요한 stochastic variation 또는 stochastic detail을 생성할 수 있으며, 상기 StyleGAN의 synthesis network에서는 각각의 스케일에서 noise(B)를 주입하여 stochastic variation을 구현하도록 각각의 convolution layer 이후에 noise(B, per-pixel noise)를 주입할 수 있다.

【0137】본 발명의 일 실시형태에서 제1 디지털 미디어 컨텐츠 생성 모델로서 StyleGAN에서는 각각의 convolution layer의 후단에 noise(B, per-pixel noise)를 주입할 수 있으며, noise(B)가 주입되어 stochastic variation 또는 stochastic detail의 랜덤한 생성의 자유도가 증가된 synthesis network의 데이터 흐름에 대해 AdaIN(Adaptive instant Normalization)을 적용하여 스타일이 전이(style transfer)되도록 할 수 있으며, 스타일 전이를 위한 AdaIN 후단에는 또 다른 convolution layer(Conv 3x3)가 연결되면서 convolution layer(Conv 3x3)-noise(B) 주입-AdaIN(Adaptive instant Normalization)의 synthesis network의 데이터 처리가 구현되거나 또는 업-샘플링(Upsample)을 거친 후에 convolution layer-noise(B) 주입-AdaIN(Adaptive instant Normalization)의 synthesis network의 데이터 처리가 구현될 수 있다.

【0138】본 발명의 일 실시형태에서, 원본 데이터(y, 스타일 데이터에 해당됨)로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 생성 모델로서, StyleGAN은 GAN 기반의 생성형 AI 모델로서, Generator G와 Discriminator D를 포함할 수 있으며, Generator G는 레이턴트 표현 z(레이턴트 벡터, random vector, latent code, 1x512)로부터 컨텐츠 데이터(x)에 표현된 컨텐츠를, 스타일 데이터(y, 또는 원본 데이터)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있고, 보다 구체적으로, 스타일 데이터(y, 또는 원본 데이터) 또는 스타일 데이터(y)로부터 추출된 스타일 정보(y^)로서 feature map의 mean(μ , 평균)과 standard deviation(σ , 표준편차)를 추종하는 가상 데이터를 생성할 수 있으며, 예를 들어, 본 발명의 일 실시형태에 따른 StyleGAN에서는 컨텐츠 데이터(x, 예를 들어, 1024x1024의 컨텐츠 데이터)로부터 추출된 고정된 스케일(4x4x215)의 저해상도의 feature map으로부터 업-스케일링을 통하여 최종적으로 고해상도의 이미지(예를 들어, 1024x1024)로 업 샘플링을 구현하면서 멀티-스케일을 형성하는 각각의 스케일 단계에서 스타일 데이터(y)로부터 추출된 스타일 정보(y^)를 전이(style transfer)시키는 방식으로 컨텐츠 데이터(x)에 표현된 컨텐츠를, 스타일 데이터(y)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있다. 이와 같이 본 발명의 일 실시형태에서 원본 데이터(또는 스타일 데이터 y)로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델에서는 멀티-스케일을 형성하는 각각의 스케일 단계에서 스타일 이미지로부터 추출된 coarse style, middle style, fine style을 synthesis network의 생성 데이터로 전이시킬 수 있으며, 본 발명의 일 실

시형태에서 상기 StyleGAN에는 저해상도의 이미지(또는 저해상도의 feature map)로부터 고해상도의 가상 이미지를 향하여 progressive 하게 업-스케일링으로 전개되는 멀티-스케일의 각각의 스케일에서 Generator의 생성과 Discriminator D의 판별을 통하여 학습이 이루어지는 progressive GAN(PGGAN)의 학습이 적용될 수 있으며, 예를 들어, 점진적으로 이미지를 업-스케일링시키면서 학습을 진행하여, 예를 들어, 4x4 --> 8x8 --> 16x16 --> 512x512 --> 1024x1024까지 업-스케일링을 진행하면서 학습이 진행될 수 있고, 멀티-스케일을 형성하는 각각의 스케일에서 그 하류의 스케일을 향하여 backpropagation 시키면서 모델의 파라메타를 학습시킬 수 있다.

【0139】본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모듈로서 StyleGAN은 GAN 기반의 생성형 AI 모델로서, StyleGAN의 학습을 위하여 Generator G와 Discriminator D를 포함하는 네트워크로 형성될 수 있으며, Generator G는 컨텐츠 데이터(x)에 표현된 컨텐츠를, 스타일 데이터(y, 또는 원본 데이터)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있고, 보다 구체적으로, Generator G로부터 생성되는 가상 데이터가 스타일 데이터(y 또는 원본 데이터) 또는 스타일 데이터(y)로부터 추출된 스타일 정보(y^)로서 feature map의 mean(μ , 평균)과 standard deviation(σ , 표준편차)을 추종하도록 Generator G로부터 생성되는 가상 데이터 또는 가상 데이터로부터 추출된 스타일 정보로서 feature map의 mean(μ , 평균)과 standard deviation(σ , 표준편차) 사이에서 산출되는 style loss(Ls)와, 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 컨

텐츠 정보로서 feature map의 각 픽셀의 값과 Generator G로부터 생성되는 가상 데이터 또는 가상 데이터로부터 추출된 컨텐츠 정보로서 feature map의 각 픽셀의 값 사이의 거리(L2 norm)에 기반하여 산출되는 contents loss(Lc)가 조합된 loss function(L)이 최소화되도록 하여, Discriminator D가 real로 판단할 만큼 real에 가까운 가상 데이터를 생성하도록 Generator G의 파라메타가 갱신될 수 있다. 예를 들어, 본 발명의 일 실시형태에서, Generator G로부터 생성된 가상 데이터와 스타일 데이터(y) 사이의 style loss(Ls)와, Generator G로부터 생성된 가상 데이터와 컨텐츠 데이터(x) 사이의 contents loss(Lc)와, 이들 style loss와 contents loss에 스케일 팩터(λ)를 고려한 loss function(L)은 이하와 같이 예시될 수 있다. 여기서, 하기의 μ(Φ(s))는 스타일 데이터(y)로부터 추출된 feature statistics의 평균(mean)을 의미할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, μ(f(s))에 해당될 수 있으며, 하기의 σ(Φ(s))는 스타일 데이터(y)로부터 추출된 feature statistics의 표준편차(standard deviation)을 의미할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, σ(f(y))에 해당될 수 있다.

【0140】

【0141】

【0142】

【0143】예를 들어, 본 발명의 일 실시형태에서, 상기 t는 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 컨텐츠 정보로서 feature map(Convolution layer Conv 3x3을 통하여 출력된 값)을 입력으로 하여 noise(B)가 더해진 synthesis network의 데이터 흐름 상의 생성 데이터에 대해 AdaIN이 적용된 데이터(t)를 의미할 수 있으며, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 스타일 정보(y^)가 전이되도록 AdaIN으로 instance normalization된 데이터(t)를 의미할 수 있고, g(t)는 AdaIN으로 instance normalization된 데이터(t)를 입력으로 하여 Generator G로부터 생성된 가상 데이터(g(t))를 의미할 수 있으며, f(g(t))는 Generator G로부터 생성된 가상 데이터(g(t))를 입력으로 하는 Discriminator의 출력을 의미할 수 있고, Φ(g(t))는 가상 데이터(g(t))를 입력으로 하여 Discriminator의 출력을 의미할 수 있다.

【0144】예를 들어, 본 명세서에서는 도에 도시된 바와 같이 이해의 편의를 위하여 상기 Discriminator D(f,Φ)는 컨텐츠 데이터(x)의 흐름을 따라 contents loss(Lc)를 산출하는 네트워크(f)와 스타일 데이터(y)의 흐름을 따라 style loss(Ls)를 산출하는 네트워크(Φ)를 포함하는 것으로 도시하였으나, 본 발명의 다양한 실시형태에서, 상기 Discriminator는 입력된 데이터가 real(ground truth) 인지 또는 fake(Generator G로부터 생성된 가상 데이터) 인지를 식별하는 것을 목표

로 하여 학습된 다양한 아키텍처의 네트워크로 구현될 수 있다.

【0145】 예를 들어, 본 발명의 일 실시형태에서, 상기 Generator G의 목표는 Discriminator D가 real로 판단할 만큼 real(ground truth)에 가까운 가상 데이터를 생성하는 것이고, Discriminator D의 목표는 Generator G에서 생성된 가상 데이터를 분별하는 것으로, 하기와 같은 목표 내지는 목표 함수의 minmax를 달성하도록 학습될 수 있다. 아래의 D(α)는 real(ground truth) 데이터 α에 대해 Discriminator D가 real로 판단할 확률을 의미할 수 있으며, D(α^)는 Generator G로부터 생성된 데이터(fake)에 대해 Discriminator D가 real로 판단할 확률을 의미할 수 있다.

【0146】 $\min(G)\max(D) V(G, D) = E_{\alpha \sim p_r}[\log D(\alpha)] + E_{\alpha^ \sim p_g}[\log(1-D(\alpha^))]$

【0147】 예를 들어, 본 발명의 일 실시형태에서 제1 디지털 미디어 컨텐츠 생성 모델로서 GAN 기반의 StyleGAN에서는 서로 다른 네트워크로서 Generator G와 Discriminator D(f,Φ)를 학습시킬 수 있으며, 보다 구체적으로 상기 Generator G의 학습에서는 스타일 데이터(y 또는 원본 데이터) 또는 스타일 데이터(y)로부터 추출된 스타일 정보로서 feature map의 mean(μ, 평균)과 standard deviation(σ, 표준편차)를 추종하도록 Generator G로부터 생성되는 가상 데이터 또는 가상 데이터로부터 추출된 스타일 정보로서 feature map의 mean(μ, 평균)과 standard deviation(σ, 표준편차)와 사이에서 산출되는 style loss(Ls)와, 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 컨텐츠 정보로서 feature map의 각 픽셀

의 값과 Generator G로부터 생성되는 가상 데이터 또는 가상 데이터로부터 추출된 컨텐츠 정보로서 feature map의 각 픽셀의 값 사이의 거리(L1 norm 또는 L2 norm)로 산출되는 contents loss(Lc)가 조합된 loss function(L)이 최소화되도록 Generator G가 학습되거나 또는 상기 style loss(Ls)와 contents loss(Lc)가 취합된 loss function(L)으로 산출된 값에 따라 Generator G로부터 생성된 가상 데이터에 대해 상기 Discriminator D(f,Φ)로부터 real(예를 들어, ground truth에 해당되는 레이블 1) 또는 real에 가깝게 판단될 수 있도록 학습될 수 있으며, 상기 Discriminator D(f,Φ)의 학습에서는 상기 style loss(Ls)와 contents loss(Lc)가 취합된 loss function(L)으로 산출된 값에 따라 real(ground truth)에 대해 레이블 1(또는 레이블 1에 가까운 확률)을 출력하고 fake(Generator G로부터 생성된 가상 데이터)에 대해 레이블 0(또는 레이블 0에 가까운 확률)을 출력하도록 학습될 수 있다.

【0148】 본 발명의 다양한 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델로서, StyleGAN은 GAN 기반의 생성형 AI 모델(generative model)로서 Generator G와 Discriminator D의 학습에서는 앞서 설명된 바와 같은 목표 내지는 목표 함수의 minmax를 달성하도록 학습될 수 있으며, 본 발명의 다양한 실시형태에서, 상기 Generator G와 Discriminator D의 학습에서는 학습 데이터의 training distribution(학습 분포)과 생성되는 가상 데이터의 generated distribution(예측되는 생성 분포) 사이에 서로에 대해 겹쳐지는 부분이 없이 서로 다른 두 분포가 서로로부터 이격되어 있을 경우에, 이들 서로 다른 분포 사이의 거리에 기반한 손

실 함수(loss function)의 그래디언트(gradient)가 산출되지 않는 문제(예를 들어, gradient vanishing)를 고려하여 두 분포 사이의 거리를 유연하게 판단하면서도 수렴의 측면을 고려한 WGAN(Wasserstein GAN)을 적용할 수 있으며, 상기 WGAN에서는 이하와 같은 목표 내지는 목표 함수의 minmax를 달성하도록 Generator G와 Discriminator D가 학습될 수 있다.

【0149】 $\min_G \max_D V(G,D) = E_{\alpha \sim p_r}[D(\alpha)] - E_{\alpha\hat{} \sim p_g}[D(\alpha\hat{})]$

【0150】 <Style Transfer>

【0151】 도 16에는 컨텐츠 데이터(x)와 스타일 데이터(y)로부터 이들 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature(또는 feature map)를 추출하고 차원을 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 스타일 데이터(y)로부터 추출된 feature statistic로부터 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여 instant normalization 시키기 위한 AdaIN(Adaptive Instant Normalization)과, 인코더로부터 차원 축소된 feature(또는 AdaIN이 적용된 feature)를 입력으로 하여 차원 확장시키면서 고해상도의 이미지를 생성하기 위한 디코더(Decoder)를 포함하며, 학습 또는 파인-튜닝을 위한 style loss(Ls) 및 contents loss(Lc)를 산출하기 위한 loss computing 네트워크(예를 들어, VGG Encoder)를 포함하는 인코더-디코더 아키텍처의 Style Transfer 네트워크를 설명하기 위한 도면이 도시되어 있다.

【0152】 도 17에는 본 발명의 일 실시형태에서, 컨텐츠 데이터(x)와 스타일

데이터(y)를 입력으로 하여 가상 데이터를 생성하기 위한 제1 컨텐츠 미디어 생성 모델을 설명하기 위한 도면으로, 컨텐츠 데이터(x)와 스타일 데이터(y)로부터 이들 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature(또는 feature map)를 추출하고 차원을 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 스타일 데이터(y)로부터 추출된 feature statistic로부터 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여 instant normalization 시키기 위한 AdaIN(Adaptive Instant Normalization)과, 인코더로부터 차원 축소된 feature(또는 AdaIN이 적용된 feature)를 입력으로 하여 차원 확장시키면서 고해상도의 이미지를 생성하기 위한 디코더(Decoder)를 포함하는 Style Transfer 네트워크로 구현된 제1 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면이 도시되어 있다.

【0153】도 18에는 본 발명의 일 실시형태에서 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델을 설명하기 위한 도면으로, 컨텐츠 데이터(x)와 스타일 데이터(y)로부터 이들 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature(또는 feature map)를 추출하고 차원을 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 스타일 데이터(y)로부터 추출된 feature statistic로부터 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여 instant normalization 시키기 위한 AdaIN(Adaptive Instant Normalization)과, 인

코더로부터 차원 축소된 feature(또는 AdaIN이 적용된 feature)를 입력으로 하여 차원 확장시키면서 고해상도의 이미지를 생성하기 위한 디코더(Decoder)를 포함하고, 학습 또는 파인-튜닝을 위한 style loss(Ls) 및 contents loss(Lc)를 산출하기 위한 loss computing 네트워크(f,ϕ)를 포함하는 Style Transfer 네트워크를 설명하기 위한 도면이 도시되어 있다.

【0154】본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델은, 인코더-디코더(Encoder-Decoder) 구조의 아키텍처를 포함하는 Style transfer 네트워크로 구현될 수 있다. 상기 Style transfer 네트워크는 컨텐츠 데이터(x)와 스타일 데이터(y, 예를 들어, 원본 데이터)를 입력으로 하여, 이들 컨텐츠 데이터(x) 및 스타일 데이터(y, 예를 들어, 원본 데이터)로부터 feature(또는 feature map)를 추출하고 차원의 축소시키기 위한 인코더(Encoder, 예를 들어, VGG Encoder)와, 상기 인코더(Encoder)로부터 출력되는 컨텐츠 데이터(또는 컨텐츠 데이터로부터 차원 축소된 feature map)에 대해 instant normalization을 수행하되, 상기 인코더(Encoder)로부터 출력되는 스타일 데이터(y 또는 스타일 데이터로부터 차원 축소된 feature map)로부터 추출된 스타일(예를 들어 feature statistics)이 컨텐츠 데이터(x)로 전이되도록, 스케일링(scaling) 파라메타 및 바이어스(bias) 파라메타를 추출하고 추출된 스케일링 파라메타와 바이어스 파라메타를 적용하여 instant normalization을 구현하기 위한 AdaIN(Adaptive Instant Normalization)과, 스타일 데이터(y, 예를 들어, 원본 데이터) 또는 스타일 데이터(y)로부터 차원 축소된 feature(또는 feature map)로부터

산출된 스케일링 파라메타 및 바이어스 파라메타로부터 instant normalization(AdaIN)된 컨텐츠 데이터(t)로서, 예를 들어, 인코더(Encoder)를 통하여 차원 축소된 feature(또는 feature map)를 입력으로 차원 확장시키면서 고해상도의 가상 데이터(예를 들어, 가상 데이터로서 가상 이미지)를 생성하기 위한 디코더(Decoder)를 포함할 수 있다. 또한, 본 발명의 일 실시형태에서 상기 제1 디지털 미디어 컨텐츠 생성 모델로서의 Style transfer 네트워크는 Style transfer로부터 생성된 가상 데이터와 컨텐츠 데이터(x) 사이의 contents loss(예를 들어, feature의 화소 단위로 산출되는 contents loss)와, Style transfer 네트워크로부터 생성된 가상 데이터와 스타일 데이터(y) 사이의 style loss(예를 들어, feature의 statistics로부터 산출되는 style loss)를 산출하기 위한 loss computing 네트워크(예를 들어, VGG Encoder)를 포함할 수 있으며, 본 발명의 일 실시형태에서는 loss computing 네트워크로서 인코더와 동일하게 구현된 VGG Encoder를 포함할 수 있다.

【0155】본 발명의 일 실시형태에서 상기 제1 디지털 미디어 컨텐츠 생성 모델로서의 Style transfer 네트워크는 컨텐츠 데이터(x)에 표현된 컨텐츠를, 스타일 데이터(y)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있으며, 보다 구체적으로 Style transfer 네트워크로부터 생성되는 가상 데이터가 컨텐츠 데이터(x)로부터 추출된 컨텐츠 정보와 스타일 데이터(y)로부터 추출된 스타일 정보를 추종하도록, Style transfer 네트워크로부터 생성되는 가상 데이터와 컨텐츠 데이터(x) 사이에서 산출되는 contents loss(Lc)와 Style transfer 네트워크로부터 생성

되는 가상 데이터와 스타일 데이터(y) 사이에서 산출되는 style loss(Ls)가 취합된 loss function(L)을 최소화시키도록, 상기 Style transfer 네트워크의 파라메타를 학습시킬 수 있다. 보다 구체적으로, 본 발명의 일 실시형태에서 상기 style loss(Ls)는 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature(또는 feature map)의 feature statistics(예를 들어, mean μ과 standard variation σ)과 Style transfer 네트워크로부터 생성된 가상 데이터 또는 가상 데이터로부터 추출된 feature(또는 feature map)의 feature statistics(예를 들어, mean μ과 standard variation σ) 사이의 거리(예를 들어, 각각의 feature로부터 산출된 feature statistics의 mean μ끼리의 L2 norm과 각각의 feature로부터 산출된 feature statistics의 standard deviation σ끼리의 L2 norm의 합산)에 기반하여 산출될 수 있으며, 상기 contents loss는 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature(또는 feature map)의 각각의 픽셀 값과 Style transfer 네트워크로부터 생성된 가상 데이터 또는 가상 데이터로부터 추출된 feature(또는 feature map)의 각각의 픽셀 값 사이의 거리(Euclidean distance, 예를 들어, L2 norm)에 기반하여 산출될 수 있다. 예를 들어, 본 발명의 일 실시형태에서 contents loss(Lc)와 style loss(Ls) 및 이들 contents loss(Lc)와 style loss(Ls)에 스케일 팩터(λ)를 고려한 loss function(L)은 이하와 같이 예시될 수 있다. 여기서, 하기의 μ(Φ(s))는 스타일 데이터(y)로부터 추출된 feature statistics의 평균(mean)을 의미할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, μ(f(s))에 해당될 수 있으며, 하기의 σ(Φ(s))는 스타일 데이터(y)로부터 추출된

feature statistics의 표준편차(standard deviation)을 의미할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, σ(f(y))에 해당될 수 있다.

【0156】

【0157】

【0158】

【0159】 예를 들어, 본 발명의 일 실시형태에서, 상기 t는 컨텐츠 데이터(x)를 입력으로 하여 AdaIN으로 instant normalization된 데이터(t)를 의미할 수 있으며, 상기 g(t)는 디코더(decoder)로부터 생성된 가상 데이터를 의미할 수 있으며, f(g(t))는 Generator G로부터 생성된 가상 데이터(g(t))를 입력으로 하는 loss computing 네트워크의 출력(예를 들어, contents loss를 산출하기 위한 loss computing 네트워크 f의 출력)을 의미할 수 있으며, ϕ(g(t))는 가상 데이터(g(t))를 입력으로 하는 loss computing 네트워크의 출력(예를 들어, style loss를 산출하기 위한 loss computing 네트워크 ϕ의 출력)을 의미할 수 있다.

【0160】 본 발명의 일 실시형태에서 컨텐츠 데이터(x)와 스타일 데이터(y)를 입력으로 하여 차원 축소된 각각의 feature를 추출하기 위한 인코더(VGG Encoder)는 VGG net 기반의 Autoencoder(예를 들어, CNN 계열의 VGG net의 layer를 포함)로

서 사전에 학습될 수 있으며, 예를 들어, 상기 AdaIN(Adaptive Instant Normalization)은 인코더(Encoder, VGG Encoder)로부터 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 추출되는 feature(또는 feature map)로부터 연산을 통하여 산출될 수 있으므로, 이들 인코더와 AdaIN은 네트워크의 학습 대상에 해당되지 않을 수 있다. 예를 들어, 본 발명의 일 실시형태에서, Style transfer 네트워크의 학습 대상은 AdaIN으로 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 고해상도의 가상 데이터를 생성하기 위한 디코더(decoder)에 해당될 수 있으며, 본 발명의 일 실시형태에서, 상기 loss computing 네트워크는 인코더(Encoder)와 동일하게 구현된 VGG Encoder로 제공될 수 있으며, 본 발명의 다양한 실시형태에서 상기 loss computing 네트워크도 디코더와 함께 학습 대상에 해당될 수도 있다.

【0161】예를 들어, 본 발명의 일 실시형태에서 상기 제1 디지털 미디어 컨텐츠 생성 모델로서 Style transfer는 Generator G와 Discriminator D를 포함하는 GAN 기반의 아키텍처로 이해될 수 있으며, 예를 들어, AdaIN의 후단에 연결된 디코더(g)를 AdaIN으로 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여, 가상 데이터의 생성을 위한 Generator G로 이해할 수 있으며, Generator G(디코더 g)로부터 생성된 가상 데이터로부터 contents loss(Lc)를 산출하기 위한 네트워크(f)와 style loss(Ls)를 산출하기 위한 네트워크(Φ)를 포함하는 loss computing 네트워크를 Discriminator D로 이해할 수 있으며, 본 발명의 일 실시형태에서 상기 Discriminator D로서 contents loss(Lc)를 산출하기 위한 네트워크(f) 및/또는 style loss(Ls)를 산출하기 위한 네트워크(Φ)는 도에 도시된 바와 같이 인코더

(Encoder, VGG Encoder)와 동일하게 구현되거나 또는 GAN에서와 유사하게 Discriminator D가 real(ground truth)과 fake(생성된 가상 데이터)를 식별할 수 있도록 학습될 수도 있으며, 예를 들어, 상기 contents loss와 style loss를 취합한 loss function으로부터 산출된 값에 따라, real(ground truth)에 대해서는 레이블 1(레이블 1에 가까운 확률)을 출력하고, fake(생성된 가상 데이터)에 대해서는 레이블 0(레이블 0에 가까운 확률)을 출력하도록 학습될 수 있다.

【0162】 예를 들어, 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델로서 Style transfer 네트워크에서는, 디코더(decoder) 또는 Generator G(g)와 Discriminator D(f,ϕ)의 서로 다른 네트워크를 학습시킬 수 있으며, 예를 들어, 상기 디코더(decoder) 또는 Generator G(g)의 학습에서는 Style transfer로부터 생성된 가상 데이터와 컨텐츠 데이터(x) 사이의 contents loss(예를 들어, 각각의 가상 데이터 및 컨텐츠 데이터로부터 추출된 feature로부터 픽셀 단위로 산출되는 contents loss)와 Style transfer로부터 생성된 가상 데이터와 스타일 데이터(y) 사이의 style loss(예를 들어, 각각의 가상 데이터 및 컨텐츠 데이터로부터 추출된 feature의 feature statistics로부터 산출되는 style loss)가 스케일 팩터(λ)를 적용하여 합산된 loss function을 최소화시키도록 학습될 수 있으며, 상기 loss computing 네트워크의 학습에서는 상기 contents loss와 style loss가 취합된 loss function의 값에 따라, real(ground truth)과 fake(가상 데이터)를 서로 식별할 수 있도록 real(ground truth)에 대해서는 레이블 1(또는 레이블 1에 가까운 확률)로 판단하고 fake(가상 데이터)에 대해서는 레이블 0(또는 레이블 0에

가까운 확률)로 판단할 수 있도록 학습될 수 있다.

【0163】<가상 데이터의 생성을 위한 데이터 증강 및 원본 데이터와 가상 데이터를 학습 데이터로 하는 파인-튜닝>

【0164】본 발명의 일 실시형태에 따른 데이터 증강(data augmentation)에 기반한 디지털 미디어 컨텐츠 생성 시스템은, 데이터 수집의 어려움과 같은 희귀성의 제한으로 한정된 개수의 원본 데이터로부터 원본 데이터의 스타일 또는 분포(예를 들어, 원본 데이터로부터 추출된 feature의 statistics)를 추종하는 가상 데이터를 생성하여 데이터 증강을 구현하는 제1 디지털 미디어 컨텐츠 생성 모델과, 한정된 개수의 원본 데이터와 제1 디지털 미디어 컨텐츠 생성 모델로부터 생성된 가상 데이터를 취합한 학습 데이터로부터 파인-튜닝(fine tuning)된 파라메타를 포함하는 생성형 AI 모델로서 생성 조건(예를 들어, 합성 데이터의 생성을 가이드 하기 위한 생성 조건)을 주입하여 생성 조건을 추종하는 합성 데이터를 생성하기 위한 제2 디지털 미디어 컨텐츠 생성 모델을 포함할 수 있다.

【0165】본 발명의 일 실시형태에서, 상기 제1 디지털 미디어 컨텐츠 생성 모델은, 사전에 대규모의 학습 데이터로 학습된 training된 모델로서 원본 데이터로부터 가상 데이터를 생성하기 위한 생성형 AI 모델에 해당될 수 있으며, 상기 제2 디지털 미디어 컨텐츠 생성 모델은 원본 데이터 및 가상 데이터가 취합된 학습 데이터로부터 파인-튜닝을 위한 생성형 AI 모델에 해당될 수 있다.

【0166】<Diffusion 기반 데이터 증강 및 파인-튜닝>

【0167】본 발명의 일 실시형태에서, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은 앞서 설명된 바와 같은 Diffusion 모델로 구현될 수 있으며, 예를 들어, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은, 실질적으로 동일 유사한 아키텍처를 포함하는 Diffusion 모델로 구현되되, 제1 디지털 미디어 컨텐츠 모델은 파라메타의 갱신을 위한 학습(training) 또는 파인-튜닝(fine tuning)에서는 요구되되 추론(inference)에서는 필요하지 않을 수 있는 forward process(diffusion process)를 포함하지 않거나 또는 forward process(diffusion process)를 수행하지 않을 수 있으며, 생성 내지는 추론(inference)을 위한 reverse process(denoising process)와 가상 데이터의 생성 조건(예를 들어, 텍스트 conditioning 또는 이미지 conditioning)을 주입하기 위한 인코더를 포함하여 reverse process(denoising process)와 생성 조건으로 주입된 텍스트 conditioning 또는 이미지 conditioning을 텍스트 임베딩 또는 이미지 임베딩으로 인코딩하여 생성 조건으로 주입하기 위한 임베딩 내지는 인코딩을 수행할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델은, 원본 데이터를 이미지 conditioning으로 주입하고 원본 데이터에 관한 변형 정보 또는 가상 데이터에 표현될 컨텐츠 정보를 텍스트 conditioning으로 주입할 수 있으며, 이를 위하여, 상기 인코더는 가상 데이터의 생성 조건으로서 원본 데이터를 포함하는 이미지 conditioning의 주입을 위한 것으로, 입력된 원본 데이터를 이미지 임베딩으로 인코딩하기 위한 이미지 인코더와, 원본 데이터에 대한 변형 정보 또는 가상 데이터에 표현될 컨텐츠 정보를 포함하는 텍스트 conditioning

의 주입을 위한 것으로, 입력된 원본 데이터에 대한 변형 정보 또는 가상 데이터로 표현될 컨텐츠 정보를 텍스트 임베딩으로 인코딩하기 위한 텍스트 인코더를 포함할 수 있으며, 상기 텍스트 인코더 및 이미지 인코더는 텍스트-이미지의 멀티 모달의 임베딩 공간을 학습한 멀티 모달의 AI 모델일 수 있으며, 예를 들어, conditioning으로 주입된 원본 데이터와 원본 데이터에 대한 변형 정보 또는 가상 데이터로 표현될 컨텐츠 정보를 각각 임베딩으로 변환한 후에 이미지 임베딩과 가장 가까운 텍스트 임베딩을 예측하거나 또는 반대로 텍스트 임베딩과 가장 가까운 텍스트 임베딩을 예측할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, 이미지-텍스트의 쌍을 서로 연결하도록 학습된 멀티 모달의 AI 모델로서, CLIP(contrastive language image pre-training)이 적용될 수 있으며, 예를 들어, 본 발명의 일 실시형태에서 상기 이미지 인코더 및 텍스트 인코더는 각각 입력된 원본 데이터를 클립 이미지 임베딩(CLIP image embedding)으로 인코딩하거나 또는 입력된 원본 데이터에 대한 변형 정보 또는 가상 데이터에 표현될 컨텐츠 정보를 클립 텍스트 임베딩(CLIP text embedding)으로 인코딩할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 텍스트 인코더는 이미지 임베딩과 매칭될 수 있는 텍스트 임베딩을 생성할 수 있으며, 텍스트 인코더로부터의 텍스트 임베딩은 생성 대상 이미지와의 크로스 어텐션(cross-attention)을 통하여 가상 데이터의 생성에 대한 conditioning을 주입할 수 있으며, 보다 구체적으로 생성 대상 이미지의 데이터(레이턴트 표현)를 쿼리(query)로 하고, 입력된 텍스트 임베딩을 키(key)와 밸류(value)로 하는 크로스 어텐션(cross attention)을 수행할 수 있다.

【0168】도 9를 참조하면, 본 발명의 일 실시형태에서, 상기 멀티 모달의 AI 모델은, 이미지와 텍스트를 각각 임베딩으로 변환한 후 이미지 임베딩과 가장 가까운 텍스트 임베딩을 예측하거나 또는 반대로 텍스트 임베딩과 가장 가까운 텍스트 임베딩을 예측할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, 이미지-텍스트의 쌍을 서로 연결하도록 학습된 멀티 모달의 AI 모델로서, CLIP(contrastive language image pre-training)을 적용할 수 있으며, 본 발명의 일 실시형태에서 상기 멀티 모달의 AI 모델은, 웹 기반의 로-데이터(raw data), 그러니까, 인간으로부터 human annotation이 들어가지 않은 로 대규모의 로-데이터(raw data)를 학습 데이터 세트로 할 수 있으며, 예를 들어, 웹 기반의 이미지와 상기 이미지를 설명하는 captioning 텍스트를 이미지-텍스트의 쌍(image-text pair)으로 하는 학습 데이터로 하여 학습될 수 있으며, contrastive learning을 이용하여 학습될 수 있다.

【0169】상기 멀티 모달의 AI 모델은, 입력된 이미지에 관한 이미지 표현(visual representation)을 산출하기 위한 이미지 인코더와 입력된 텍스트에 관한 텍스트 표현(language representation) 표현을 산출하기 위한 텍스트 인코더를 포함할 수 있으며, 서로 n개의 쌍을 형성하는 이미지-텍스트(image-text pair)의 쌍을 각각의 이미지 인코더와 텍스트 인코더에 입력하여, 이미지의 미니 배치(mini-batch, n개의 이미지 학습 데이터)와 텍스트의 미니 배치(mini-batch, n개의 텍스트 학습 데이터)를 생성한 후에, 서로 쌍을 이루는 텍스트-이미지 쌍을 파지티브 쌍(positive pair)으로 하고, 서로 쌍을 이루지 않는 텍스트-이미지 쌍은 네거티브 쌍(negative pair)로 하여, n개의 파지티브 쌍(positive pair)에 대한 코사인 유사

도(cosine similarity)는 최대가 되고, n^2-n개의 네거티브 쌍(negative pair)에 대한 코사인 유사도(cosine similarity)는 최소가 되도록 학습될 수 있으며, 이와 같은 contrastive learning으로부터 이미지 인코더와 텍스트 인코더를 함께 학습시킬 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 멀티 모달의 AI 모델은, 이미지-텍스트의 멀티 모달(multi-modal)의 임베딩 공간을 학습할 수 있으며, 예를 들어, 입력된 이미지를 이미지 스페이스로 임베딩하는 이미지 인코더와 입력된 텍스트를 텍스트 스페이스로 임베딩하는 텍스트 인코더로부터 각각의 이미지 임베딩과 텍스트 임베딩을 서로 매핑시키도록, 상기 이미지 인코더와 텍스트 인코더를 학습시킬 수 있다.

【0170】본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 생성 모델과 달리, 원본 데이터와 가상 데이터가 취합된 학습 데이터로부터 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델은 파라메타의 갱신을 위한 학습 또는 파인-튜닝을 위한 forward process(diffusion process)를 포함하고 forward process(diffusion process)를 수행할 수 있으며, 이에 따라 상기 제2 디지털 미디어 컨텐츠 생성 모델은, 생성 또는 추론을 위한 reverse process(denoising process)와 합성 데이터의 생성 조건(예를 들어, 텍스트 conditioning 또는 이미지 conditioning)을 주입하기 위한 인코더를 포함하여 생성 조건으로 주입된 텍스트 conditioning 또는 이미지 conditioning을 텍스트 임베딩 또는 이미지 임베딩으로 인코딩하여 생성 조건으로 주입하기 위한 임베딩 내지는 인코딩을 수행할 수 있으며, 상기 제2 디지털 미디어

컨텐츠 생성 모델은 생성 또는 추론을 위한 reverse process와 합성 데이터의 생성 조건을 주입하기 위한 인코더와 함께, 학습 또는 파인-튜닝을 위한 forward process를 포함하여 forward process를 수행할 수 있다. 본 발명의 일 실시형태에서, 원본 데이터 및 가상 데이터가 취합된 학습 데이터로부터 파라메타의 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 모델은 원본 데이터 및 가상 데이터가 취합된 학습 데이터로서 원본 이미지로부터 시간 스텝을 전진시키면서 노이즈 스케줄(noise schedule)에 따라 노이즈를 추가하여 점진적으로 원본 이미지의 패턴을 붕괴시키는 노이징 프로세스(noising process)와, 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복원되도록 노이지한 이미지로부터 상대적으로 원본 이미지에 가까운 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현할 수 있으며, 앞서 설명된 바와 같이, reverse process를 근사하는 신경망 모델은 현재의 시간 스텝에서의 이미지(Xt)로부터 덜 노이지(noisy)한 이전 시간 스텝에서의 이미지(Xt-1)로부터 추가된 가우시안 노이즈를 예측함으로써, 현재의 시간 스텝에서의 덜 노이지한 이미지(Xt)의 평균을 예측할 수 있으며, 신경망 모델로부터 예측된 평균과 forward process에서 산출된 실제 평균(타겟 레이블)과의 오차를 최소화시키도록 신경망 모델의 목적 함수가 구성될 수 있다.

【0171】본 발명의 일 실시형태에서 상기 제2 디지털 미디어 컨텐츠 생성 모델은, forward process와 reverse process를 수행하면서 원본 이미지의 특정한 패턴이 붕괴된 가우시안 노이즈로부터 원본 데이터의 패턴을 복원시키는 reverse process를 학습할 수 있으며, 가우시안 노이즈로부터 예측된 노이즈를 제거하면서

시간의 스텝에 따라 덜 노이지한 이미지를 생성하면서 원본 이미지의 패턴을 복원하는 reverse process를 학습할 수 있다. 이와 같이 본 발명의 일 실시형태에서, 상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 reverse process를 학습할 수 있으며, 이때, 상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 생성 조건으로서 이미지 conditioning 또는 텍스트 conditioning은 주입하지 않을 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 이미지 conditioning 및 텍스트 conditioning의 주입을 위한 이미지 인코더 또는 텍스트 인코더로부터의 이미지 임베딩 또는 텍스트 임베딩은, 생성 대상 이미지의 데이터(레이턴트 표현)를 쿼리(query)로 하고, 입력된 이미지 임베딩 또는 텍스트 임베딩을 키(key)와 밸류(value)로 하는 크로스 어텐션을 수행할 수 있으며, 이때, 크로스 어텐션을 산출하기 위한 퀴리 행렬, 키 행렬, 밸류 행렬은 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝 이전에 학습되어 있을 수 있으며, 예를 들어, 가상 데이터의 생성 조건으로서 제1 디지털 미디어 컨텐츠 생성 모델은, 대규모의 학습 데이터로부터 앞서 설명된 reverse process와 가상 데이터의 생성 조건을 주입하기 위한 이미지 conditioning과 텍스트 conditioning의 주입을 위한 이미지 인코더, 텍스트 인코더 및 크로스 어텐션에 대해 학습되어 있는 training된 모델에 해당될 수 있으며, 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델을 통하여 원본 데이터로부터 생성된 가상 데이터와, 원본 데이터를 취합한 학습 데이터로부터의 파인-튜닝에서는, 원본 데이터와 원본 데이터로부터 스타일이 전이(style transfer)되어 동일 또는 유사한 스타일로 표현된 가상 데이터로부터, 그러니까, 서로 동일 또는

유사한 스타일로 표현된 학습 데이터(원본 데이터와 가상 데이터)로부터 이미지 상에 표현되어 있는 컨텐츠(예를 들어, 이미지 상에 표현되어 있는 주제 또는 객체)가 서로 다른 합성 데이터를 생성하기 위한 것으로, 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 서로 동일 또는 유사한 스타일의 이미지를 생성하기 위한 reverse process를 학습하는 것으로 충분할 수 있으며, 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서 스타일 전이를 위한 학습 데이터로서 이미지 conditioning과 생성 조건으로서 학습 데이터의 변형 조건 또는 합성 데이터로 표현될 컨텐츠로서 텍스트 conditioning을 주입하더라도, 생성되는 또는 예측되는 합성 데이터가 추종할 타켓 데이터(타켓 이미지)를 입수할 수 없을 수 있다는 점에서, 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 동일 또는 유사한 스타일의 합성 데이터의 생성을 위한 reverse process는 학습하되, 이미지 conditioning 및 텍스트 conditioning은 주입하지 않을 수 있다.

【0172】<StyleGAN 기반 데이터 증강 및 파인-튜닝>

【0173】본 발명의 일 실시형태에서, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은 앞서 설명된 바와 같은 StyleGAN 모델로 구현될 수 있으며, 예를 들어, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은, 실질적으로 동일 유사한 아키텍처를 포함하는 StyleGAN 모델로 구현되되, 제1 디지털 미디어 컨텐츠 모델은 파라메타의 갱신을 위한 학습(training) 또는 파인-튜닝(fine tuning)에서는 요구되되 추론(inference)에서는 필요하지 않을 수 있는 Discriminator(f,Φ)를 포함하지 않거나 또는 Discriminator(f,Φ)의 프로세스를 수행하지 않을 수 있으며, 생성 내지는 추

론(inference)을 위한 Generator G의 프로세스를 수행할 수 있다.

【0174】 예를 들어, 본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델은, 원본 데이터로서 스타일 데이터(y)와 컨텐츠 데이터를 입력하여 컨텐츠 데이터(x)에 표현된 컨텐츠(주제 또는 객체)를, 스타일 데이터(y)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있으며, 예를 들어, 이와 같은 가상 데이터의 생성 또는 추론에서는 컨텐츠 데이터와 가상 데이터 사이의 contest loss와 스타일 데이터와 가상 데이터 사이의 style loss를 산출할 필요가 없으므로, 이들 contents loss(Lc)와 style loss(Ls)가 취합된 loss function(L)의 산출을 위한 Discriminator 네트워크 또는 loss computing 네트워크가 요구되지 않을 수 있다.

【0175】 원본 데이터로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 생성 모델과 달리, 원본 데이터와 가상 데이터가 취합된 학습 데이터로부터 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델은 파라메타의 갱신을 위한 학습 또는 파인-튜닝을 위하여 가상 데이터와 컨텐츠 데이터 사이의 contents loss(Lc) 및 가상 데이터와 스타일 데이터 사이의 style loss(Ls)가 취합된 loss function(L)의 산출을 위한 Discriminator(f,Φ) 내지는 loss computing 네트워크를 포함하여 Discriminator 프로세스 내지는 loss computing 프로세스를 수행할 수 있으며, 예를 들어, 상기 Discriminator 프로세스 내지는 loss computing 프로세스로부터 산출된 contents loss(Lc, 예를 들어, 가상 데이터와 컨텐츠 데이터로부터 추출된 feature로부터 픽셀 단위로 산출되는 contents loss)와 style loss(Ls, 예

를 들어, 가상 데이터와 스타일 데이터로부터 추출된 feature statistics로부터 산출되는 style loss)가 취합된 loss function(L)이 최소화되도록 Generator G의 학습이 구현될 수 있다. 이와 같이, 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델을 통하여 원본 데이터로부터 생성된 가상 데이터와 원본 데이터를 취합한 학습 데이터로부터 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델은 loss computing 네트워크(f,Φ)를 포함할 수 있다.

【0176】상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 컨텐츠 데이터(x)와 스타일 데이터(y)를 입력으로 하여, 컨텐츠 데이터에 표현된 컨텐츠(주제 또는 객체)를, 스타일 데이터(y)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있으며, 저해상도 이미지(또는 저해상도 feature map)으로부터 고해상도의 가상 이미지를 향하여 progressive하게 업 스케일링으로 전개되는 멀티-스케일의 각각의 스케일에서 컨텐츠 데이터(x, 예를 들어, 1024x1024 고해상도의 컨텐츠 데이터로부터 일정한 스케일- 4x4x512의 고정된 스케일의 constant tensor) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 스타일 정보(y^)로부터 추출된 feature statistics(예를 들어, mean μ와 standard deviation σ)를 스케일링(scaling) 파라메타 및 바이어스 파라메타로 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 Generator G(예를 들어, convolution layer Conv 3x3 및/또는 Upsample)로부터 생성된 feature(g(t))를 입력으로 하는 Discriminator의 출력(f(g(t))로부터 contents loss(Lc)와 style loss(Ls)가 취합된 loss function을 산

출할 수 있으며, 산출된 loss function을 최소화시키도록 상기 StyleGAN의 네트워크를 형성하는 Generator G(예를 들어, convolution layer Conv 3x3 및/또는 Upsample)의 파라메타를 학습시킬 수 있다. 예를 들어, 본 발명의 일 실시형태에서, StyleGAN 네트워크의 파라메타의 갱신을 위한 학습 또는 파인-튜닝의 학습 대상은, progressive 하게 업 스케일링으로 전개되는 멀티 스케일을 형성하는 각각의 스케일에서 convolution layer(Conv 3x3)와 Upsample에 해당될 수 있으며, 스타일 데이터 또는 스타일 데이터로부터 추출된 feature statistics로부터 스케일링(scaling) 파라메타 및 바이어스(bias) 파라메타를 적용하는 AdaIN에 대한 학습은 필요하지 않을 수 있고, 스타일 데이터 또는 스타일 데이터로부터 추출된 feature statistics(예를 들어, mean μ, standard deviation σ)로부터 해당되는 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해 instant normalization을 구현하는 AdaIN에 대해서는 학습이 필요하지 않을 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 학습 대상으로서 convolution layer(Conv 3x3)와 Upsample은 GAN 기반의 StyleGAN에서 Generator G(예를 들어, progressive 하게 업 스케일링되는 멀티 스케일의 각각의 스케일에서의 Generator G)에 해당될 수 있고, Generator G와 Discriminator D의 학습은 멀티 스케일을 형성하는 각각의 스케일 마다 구현될 수 있다. 본 발명의 일 실시형태에서 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델로서 GAN 기반의 StyleGAN에서 Generator G(예를 들어, convolution layer Conv 3x3 및/또는 Upsample)은 학습 대상에 해당될 수 있으며, 본 발명의 다양한 실시형태에서 구체적인 StyleGAN의 구현 형태에 따라

Discriminator D는 학습 대상에 해당되거나 또는 학습 대상에 해당되지 않을 수도 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 Discriminator D는 Generator G를 형성하는 convolution layer(Conv 3x3) 및 Upsample 네트워크의 파라메타를 학습시키기 위한 loss function(예를 들어, 스케일 팩터 λ를 적용하여 가중 합산한 contents loss Lc와 style loss Ls가 취합된 loss function)를 산출하기 위한 loss computing 네트워크로서, 예를 들어, Generator G의 역-프로세스로서 상정될 수 있으며, 예를 들어, Generator G의 역-프로세스로서 예를 들어, Generator G로서 convolution layer에 대한 역-프로세스로서 transposed convolution 등의 네트워크로 상정되거나 및/또는 Generator G로서 Upsample에 대한 역-프로세스로서 Downsampling 네트워크로 상정될 수도 있다. 즉, 본 발명의 일 실시형태에서, 상기 Discriminator D는 Generator G의 파라메타를 파인-튜닝하기 위한 loss function(L)을 산출하기 위한 loss computing 네트워크로서, 예를 들어, Generator G로부터 생성된 가상 데이터 또는 가상 데이터의 feature(g(t))에 대해 역-프로세스를 적용하여 당초의 스타일 데이터(y)로부터 추출된 feature statics를 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)로 복원될 수 있는지와, Generator G로부터 생성된 가상 데이터 또는 가상 데이터의 feature(g(t))에 대해 역-프로세스를 적용하여 가상 데이터 또는 가상 데이터의 feature(g(t))의 statistics가 당초의 스타일 데이터의 feature statistics를 추종하는지를 평가하는 것으로 이해될 수 있으며, 예를 들어, Generator G로부터 생성된 가상 데이터 또는 가상 데이터의 feature(g(t))에 대해 Generator G의 역-프로세스에 해당되는

Discriminator D를 적용하여 Discriminator D의 출력(f(g(t))과 당초의 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t) 사이에서 픽셀 단위의 거리(예를 들어, L2 norm)에 기반한 contents loss(Lc)와 Discriminator D의 출력(f(g(t) 또는 ϕ(g(t))과 AdaIN에 적용된 스타일 데이터로부터 각각 추출된 feature statistics(예를 들어, 각각의 mean μ과 standard deviation σ) 사이의 거리(예를 들어, L2 norm)에 기반한 style loss(Ls)를 취합한 loss function(L)을 산출할 수 있으며 산출된 loss function(L)을 최소화시키도록, 상기 Generator 네트워크를 형성하는 convolution layer(Conv 3x3) 및 Upsample 네트워크의 파라메타를 학습시킬 수 있다.

【0177】 이와 같이 본 발명의 일 실시형태에서 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델로서 StyleGAN은 GAN 기반의 생성형 AI 모델로서 Generator G에 대한 학습은 구현하되, loss function의 산출을 위한 loss computing 네트워크로서 Discriminator D에 대한 학습은 구현하지 않을 수 있으며, 이에 따라 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는, 원본 데이터와 제1 디지털 미디어 컨텐츠 생성 모델을 통하여 원본 데이터로부터 생성된 가상 데이터로서 서로 동일 유사한 스타일로 표현된 파인-튜닝을 위한 학습 데이터(원본 데이터+가상 데이터)로서 스타일 데이터(y)와 컨텐츠 데이터(x)를 입력으로 하여 파인-튜닝이 구현될 수 있으며, 예를 들어, 별도의 타겟 레이블(예를 들어, 타겟 레이블로서 StyleGAN으로부터 생성된 가상 데이터가 추종할 목표로서의 타겟 이미지)가 필요하지 않을 수 있으며, 예를 들어, 파인-튜닝

을 위한 학습 데이터(원본 데이터+가상 데이터)로서 동일 유사한 스타일로 표현된 스타일 데이터(y)와 컨텐츠 데이터(x)를 입력으로 하여 가상 데이터가 생성되는 프로세스 상에서 상기 StyleGAN 네트워크의 파라메타가 학습될 수 있다. 예를 들어, 본 발명의 일 실시형태에서 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델로서, 앞서 설명된 Diffusion 모델(예를 들어, stable Diffusion 모델)로 구현된 제2 디지털 미디어 컨텐츠 생성 모델에서는 학습 데이터로서 동일 유사한 스타일로 표현된 파인-튜닝을 위한 학습 데이터(원본 데이터+가상 데이터)를 이미지 conditioning으로 주입하고 학습 데이터의 변형 정보 또는 가상 데이터로 표현될 컨텐츠 정보(주제 또는 객체)를 텍스트 conditioning으로 주입하는 합성 데이터의 생성 조건에 대해서는 학습하지 않고, 예를 들어, 학습 데이터의 특정한 패턴이 붕괴된 가우시안 노이즈로부터 학습 데이터의 패턴이 복원된 이미지를 생성하기 위한 reverse process(denoising process)만을 학습하였으며, 예를 들어, 상기 Diffusion 모델 기반의 제2 디지털 컨텐츠 생성 모델에서는 학습 데이터와 함께 부여될 필요가 있는 타겟 레이블(예를 들어, 네트워크로부터 예측되는 생성 이미지가 추종할 목표로서의 타겟 이미지)를 수집하기 어렵다는 점을 고려하여 이미지 생성의 조건을 부여하지 않고 학습 데이터에 대한 forward process(noising process)와 역-프로세스로서 reverse process(denoising process)를 구현하면서 당초의 학습 데이터의 패턴이 복원되는 reverse process만을 학습할 수 있다는 점에서, 예를 들어, 본 발명의 일 실시형태에서, StyleGAN 기반의 제2 디지털 미디어 컨텐츠 생성 모델에서는 학습 데이터(원본 데이터+가상 데이터)로서의 스타일 데이터와 컨텐츠

데이터를 입력으로 하여 가상 데이터의 생성을 구현하면서 네트워크의 파라메타가 학습될 수 있다는 점에서, 앞서 설명된 바와 같은 Diffusion 기반의 제2 디지털 미디어 컨텐츠 생성 모델과는 차별될 수 있다.

【0178】이와 같이, 본 발명의 일 실시형태에서, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은 앞서 설명된 바와 같은 GAN 기반의 StyleGAN 모델로 구현될 수 있으며, 예를 들어, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은, 실질적으로 동일 유사한 아키텍처를 포함하는 StyleGAN 모델로 구현되되, 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 모델은 파라메타의 갱신을 위한 학습(training) 또는 파인-튜닝(fine tuning)에서는 요구되되 추론(inference)에서는 필요하지 않을 수 있는 loss computing 네트워크(f,ϕ) 또는 Discriminator D(f,ϕ)를 포함하지 않거나 또는 loss computing 프로세스를 수행하지 않을 수 있으며, 이와 같이, 원본 데이터로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 생성 모델과 달리, 원본 데이터와 가상 데이터가 취합된 학습 데이터로부터 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델은 파라메타의 갱신을 위한 학습 또는 파인-튜닝을 위한 loss computing 네트워크(f,ϕ) 또는 Discriminator D(f,ϕ)를 포함하여 loss computing 프로세스를 수행할 수 있으며, 예를 들어, StyleGAN 기반의 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 생성 이미지(합성 이미지)의 생성 조건으로서, 예를 들어, 생성 이미지(합성 이미지)에 표현될 컨텐츠(주제 또는 객체) 정보의 주입을 위한 컨텐츠 데이터를 포함하는 생성 조건이 주입될 수 있다.

【0179】<Style transfer 기반 데이터 증강 및 파인-튜닝>

【0180】본 발명의 일 실시형태에서, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은 앞서 설명된 바와 같은 Style transfer 네트워크로 구현될 수 있으며, 예를 들어, 상기 제1, 제2 디지털 미디어 컨텐츠 모델은, 실질적으로 동일 유사한 아키텍처를 포함하는 Style transfer 네트워크로 구현되되, 제1 디지털 미디어 컨텐츠 모델은 파라메타의 갱신을 위한 학습(training) 또는 파인-튜닝(fine tuning)에서는 요구되되 추론(inference)에서는 필요하지 않을 수 있는 loss computing 네트워크(f,Φ)를 포함하지 않고, loss computing 프로세스를 수행하지 않을 수 있으며, 생성 내지는 추론(inference)을 위한 Generator G(g, 예를 들어, Decoder)의 프로세스를 수행할 수 있다.

【0181】예를 들어, 본 발명의 일 실시형태에서, 원본 데이터로부터 가상 데이터를 생성하기 위한 제1 디지털 미디어 컨텐츠 생성 모델은, 원본 데이터로서 스타일 데이터(y)와 컨텐츠 데이터를 입력하여 컨텐츠 데이터(x)에 표현된 컨텐츠(주제 또는 객체)를, 스타일 데이터(y)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있으며, 예를 들어, 이와 같은 가상 데이터의 생성 또는 추론에서는 컨텐츠 데이터와 가상 데이터 사이의 contest loss와 스타일 데이터와 가상 데이터 사이의 style loss를 산출할 필요가 없으므로, 이들 contents loss(Lc)와 style loss(Ls)가 취합된 loss function(L)의 산출을 위한 loss computing 네트워크(f,Φ)가 요구되지 않을 수 있으며, loss computing 프로세스를 수행하지 않을 수 있다.

【0182】원본 데이터로부터 가상 데이터의 생성을 위한 제1 디지털 미디어 컨텐츠 생성 모델과 달리, 원본 데이터와 가상 데이터가 취합된 학습 데이터로부터 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델은 파라메타의 갱신을 위한 학습 또는 파인-튜닝을 위하여 가상 데이터와 컨텐츠 데이터 사이의 contents loss(Lc) 및 가상 데이터와 스타일 데이터 사이의 style loss(Ls)가 취합된 loss function(L)의 산출을 위한 loss computing 네트워크(f,Φ)를 포함하여 loss computing 프로세스를 수행할 수 있으며, 예를 들어, 상기 loss computing 프로세스로부터 산출된 contents loss(Lc, 예를 들어, 가상 데이터와 컨텐츠 데이터로부터 추출된 feature로부터 픽셀 단위로 산출되는 contents loss)와 style loss(Ls, 예를 들어, 가상 데이터와 스타일 데이터로부터 추출된 feature statistics로부터 산출되는 style loss)가 취합된 loss function(L)이 최소화되도록 Generator G(g, 예를 들어, Decoder)의 학습이 구현될 수 있다. 이와 같이, 본 발명의 일 실시형태에서, 제1 디지털 미디어 컨텐츠 생성 모델을 통하여 원본 데이터로부터 생성된 가상 데이터와 원본 데이터를 취합한 학습 데이터로부터 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델은 loss computing 네트워크(f,Φ)를 포함할 수 있다.

【0183】상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는 컨텐츠 데이터(x)와 스타일 데이터(y)를 입력으로 하여, 컨텐츠 데이터에 표현된 컨텐츠(주제 또는 객체)를, 스타일 데이터(y)로부터 추출된 스타일로 표현하는 가상 데이터를 생성할 수 있으며, 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)를 입력으로 하는 인코더(Encoder, 예를 들어, VGG Encoder)를 통하여 컨텐츠 데이터(x)로부터 차

원 축소된 feature(feature map)에 대해 스타일 데이터(y) 또는 스타일 데이터(y)를 입력으로 하는 인코더(Encoder, 예를 들어, VGG Encoder)를 통하여 스타일 데이터(y)로부터 차원 축소된 feature statistics(예를 들어, mean μ와 standard deviation σ)를 스케일링(scaling) 파라메타 및 바이어스 파라메타로 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 디코더(g, Decoder)로부터 생성된 가상 데이터(g(t))를 입력으로 하는 loss computing 네트워크(f)의 출력(f(g(t))로부터 contents loss(Lc)와 style loss(Ls)가 취합된 loss function을 산출할 수 있으며, 산출된 loss function을 최소화시키도록 상기 Style transfer 네트워크를 형성하는 디코더(g)의 파라메타를 학습시킬 수 있다. 예를 들어, 본 발명의 일 실시형태에서, Style transfer 네트워크의 파라메타의 갱신을 위한 학습 또는 파인-튜닝의 학습 대상은, 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 인코더를 통하여 차원 축소된 feature(또는 feature map)에 대해, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 차원 축소된 feature statistics(예를 들어, mean μ와 standard deviation σ)로부터 추출된 스케일링(scaling) 파라메타 및 바이어스(bias) 파라메타를 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 차원 확장된 고해상도의 가상 데이터를 생성하기 위한 디코더(g)에 해당될 수 있으며, 예를 들어, 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 차원 축소된 feature를 추출하기 위한 인코더(g, Encoder, VGG Encoder)는 VGG net 기반의 Autoencoder로서, CNN 계열의 VGG net의 layer를 포함하여 사전에 학습될 수 있으며, 본 발명의 일 실시형태에서 파인-튜닝

을 위한 제2 디지털 미디어 컨텐츠 생성 모델에서 학습 내지는 파인-튜닝의 대상에 해당되지 않을 수 있다. 또한, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics로부터 스케일링(scaling) 파라메타 및 바이어스(bias) 파라메타를 적용하는 AdaIN에 대한 학습은 필요하지 않을 수 있고, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics(예를 들어, mean μ, standard deviation σ)로부터 해당되는 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature(예를 들어, 컨텐츠 데이터 x로부터 차원 축소된 feature 또는 feature map)에 대해 instant normalization을 구현하는 AdaIN에 대해서는 학습이 필요하지 않을 수 있다. 본 발명의 일 실시형태에서 파인-튜닝을 위한 제2 디지털 미디어 컨텐츠 생성 모델로서 Style transfer 네트워크에서 디코더(g)는 학습 대상에 해당될 수 있으며, 본 발명의 다양한 실시형태에서 구체적인 Style transfer 네트워크의 구현 형태에 따라 loss computing 네트워크(f,ϕ)는 학습 대상에 해당되거나 또는 학습 대상에 해당되지 않을 수도 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 loss computing 네트워크(f,ϕ)는 디코더(Decoder)의 파라메타를 학습시키기 위한 loss function(예를 들어, 스케일 팩터 λ를 적용하여 가중 합산한 contents loss Lc와 style loss Ls가 취합된 loss function)를 산출하기 위한 loss computing 네트워크로서, 예를 들어, 디코더(g)의 역-프로세스로서 상정될 수 있으며, 차원 축소를 위한 인코더(f)와 차원 확장을 위한 디코더(g)를 포함하는 인코더-디코더 네트워크에서, 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 차원 축소된 feature를 추출하기 위한 인코더(g, Encoder, VGG Encoder)로서의

VGG Encoder와 동일한 네트워크로 구현될 수 있으며, 예를 들어, VGG net 기반의 Autoencoder로서 CNN 계열의 VGG net의 layer를 포함하여 사전 학습될 수 있으며, 컨텐츠 데이터 및 스타일 데이터로부터 차원 축소된 feature를 추출하기 위한 인코더(f, Encoder)와 동일한 네트워크로 구현될 수 있으며, 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서, 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 차원 축소를 위한 인코더(f, Encoder, 예를 들어, VGG Encoder) 및 상기 인코더(f, Encoder, 예를 들어, VGG Encoder)와 동일한 네트워크로 구현될 수 있는 loss computing 네트워크(f, 예를 들어, VGG Encoder)는 학습 내지는 파인-튜닝의 대상에 해당되지 않을 수 있다. 예를 들어, 상기 loss computing 네트워크(f)는 차원 확장을 통하여 고해상도의 가상 데이터를 생성하기 위한 디코더(g, Decoder)의 역-프로세스로서 예를 들어, 차원 축소를 위한 인코더(Encoder) 네트워크로 구현될 수 있으며, 본 발명의 일 실시형태에서, 상기 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 차원 축소된 feature(또는 feature map)을 추출하기 위한 인코더(f, Encoder, VGG Encoder)와 동일한 네트워크로 구현될 수 있다.

【0184】본 발명의 일 실시형태에서, 파인-튜닝을 위한 제2 디지털 컨텐츠 생성 모델로서, StyleGAN 또는 Style transfer의 loss function의 산출을 위한 loss computing 네트워크(f,ϕ)는 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics로부터 AdaIN을 통하여 instant normalization된 컨텐츠 데이터

(t)를 입력으로 하는 Generator G(예를 들어, StyleGAN에서 convolution layer Conv 3x3 및/또는 Upsample) 또는 디코더(g, 예를 들어, Style transfer 네트워크에서 디코더 g)로부터 출력된 컨텐츠 데이터(g(t))를 입력으로 하는 loss computing 네트워크의 출력(f(g(t))과, AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)와의 픽셀 단위의 거리(L2 norm)에 기반하여 contents loss(Lc)를 산출할 수 있으며, 또한, AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)를 입력으로 하는 Generator G(예를 들어, StyleGAN에서 convolution layer Conv 3x3 및/또는 Upsample) 또는 디코더(g, 예를 들어, Style transfer 네트워크에서 디코더 g)로부터 출력된 컨텐츠 데이터(g(t))를 입력으로 하는 loss computing 네트워크의 출력(f(g(t))의 feature statistics(예를 들어, mean μ와 standard deviation σ)와, AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)의 feature statistics(예를 들어, mean μ와 standard deviation σ) 사이의 거리(예를 들어, L2 norm)에 기반하여 style loss(Ls)를 산출할 수 있다.

【0185】본 발명의 일 실시형태에서, 제2 디지털 미디어 컨텐츠 생성 모델로서, StyleGAN 또는 Style transfer의 loss computing 네트워크는, 각각의 contents loss(Lc)의 산출을 위한 네트워크(f)와 style loss(Ls)의 산출을 위한 네트워크(Φ)를 포함하는 것으로 예시될 수 있으나, 실질적으로, 상기 StyleGAN 또는 Style transfer의 loss computing 네트워크(f, Φ)는 Generator G(예를 들어, StyleGAN에서 convolution layer Conv 3x3 및/또는 Upsample) 또는 디코더(g, 예를 들어, Style transfer 네트워크에서 디코더 g)의 역-프로세스로서, 예를 들어,

StyleGAN에서 transposed convolution 및/또는 Downsample과 같은 네트워크로 구현되거나 또는 Style transfer에서 인코더(Encoder, 예를 들어, VGG Encoder)로 구현될 수 있으며, 예를 들어, 상기 loss computing 네트워크의 출력(f(g(t)) 내지는 loss computing 네트워크로부터 출력되는 feature(f(g(t))로부터 픽셀 단위의 contents loss(Lc)를 산출하거나 또는 feature statistics(예를 들어, mean μ, standard deviation σ)로부터 style loss(Ls)를 산출할 수 있으며, 이런 의미에서 본 명세서를 통하여 제2 디지털 미디어 컨텐츠 생성 모델로서 StyleGAN 또는 Style transfer의 loss computing 네트워크로서 contents loss(Lc)를 산출하는 것으로 예시된 네트워크(f)와 style loss(Ls)를 산출하는 것으로 예시된 네트워크(φ)는, 스타일 데이터(y)로부터 추출된 스케일링 파라메타 및 바이어스 파라메타를 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(x)를 입력으로 하는 loss computing 네트워크로부터 출력되는 feature로부터 픽셀 단위의 화소 값을 추출하거나 또는 mean(μ)과 standard deviation(σ)를 포함하는 feature statistics를 추출하는 후 처리 연산에 차이가 있을 뿐이고 실질적으로 동일한 네트워크로 구현되는 것으로 이해될 수 있다.

【0186】<LoRA, Low Rank Adaptation>

【0187】도 19a 및 도 19b에는 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에서, 파라메타 행렬(W)에 대한 파인-튜닝을 통한 파라메타 행렬의 조정분(ΔW)에 대해 저차원 행렬 분해(low rank decomposition)로부터 저차원 행렬인 LoRA 행렬(LoRA matrices A,B 또는 LoRA 어뎁터 A,B)의 행렬 곱의

형태로 근사시키고, 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬(W)은 고정(freeze)시키는 LoRA(Low Rank Adaptation, 도 19a)와 제2 디지털 미디어 컨텐츠 모델의 파라메타를 업-데이트 시키는 통상적인 파인-튜닝(Weight update in regular fine-tuning, 도 19b)을 설명하기 위한 도면이 각각 도시되어 있다.

【0188】 도 20에는 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬(W)에 대한 파인-튜닝을 통한 파라메타 행렬의 조정분(ΔW)에 대해 저차원 행렬 분해(low rank decomposition)로부터 저차원 행렬인 LoRA 행렬(LoRA matrices A,B 또는 LoRA 어뎁터 A,B)의 행렬 곱의 형태로 근사시키는 LoRA를 설명하기 위한 도면이 도시되어 있다.

【0189】 도 21에는 도 19a에 도시된 LoRA를 보다 구체적으로 설명하기 위한 것으로, 1) 제2 디지털 미디어 컨텐츠 모델의 파라메타 고정(freeze), 2) 파인-튜닝을 위한 학습 데이터(원본 데이터+가상 데이터)가 제2 디지털 미디어 컨텐츠 모델과 LoRA adpater(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)에 함께 입력되는 Input, 3) adapter A(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)를 통과한 학습 데이터가 adapter B(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)를 통과하면서 BA의 행렬 곱이 생성되는 Adapter Train, 4) 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬 W를 통과한 출력과 LoRA adapter를 통과한 출력이 합산되면서 최종 출력의 산출을 설명하기 위한 도면이 도시되어 있다.

【0190】 도 22에는 본 발명의 일 실시형태에서, 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에 적용되는 LoRA(Low Rank Adaptation)를 설명하기 위한 도면으

로, 이전 시간 스텝에서의 은닉 상태(Hidden state)의 서로 다른 시간 스텝에서의 출력에 대해 쿼리 행렬(W_Q), 키 행렬(W_K) 및 밸류 행렬(W_V)을 승산하여 각각의 쿼리 벡터(Q), 키 벡터(K) 및 밸류 벡터(V)를 산출하는 프로세스에서 LoRA 어뎁터가 적용되는 것을 설명하기 위한 도면이 도시되어 있다.

【0191】도 23에는 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에 적용될 수 있는 QLoRA(Quantized Low Rank Adaptation)를 설명하기 위한 도면으로, 32비트 부동 소수점(float 32, FP32)의 제2 디지털 미디어 컨텐츠 모델(Before Quantization)의 파라메타를 4비트 NormalFloat(NF4)라는 새로운 데이터 타입으로 양자화시키는(After Quantization) QLoRA(Quantized Low Rank Adaptation)를 설명하기 위한 도면이 도시되어 있다.

【0192】본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 모델에 대한 파인-튜닝에서는 LoRA(Low Rank Adaptation)를 통하여 파인-튜닝의 학습 속도를 높일 수 있으며, 예를 들어, 학습 대상이 되는 파라메타의 개수를 획기적으로 줄이면서 학습 속도 내지는 파라메타의 수렴 속도를 높일 수 있고, 파인-튜닝의 연산 자원 내지는 메모리를 절약할 수 있다. 예를 들어, 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에 적용 가능한 LoRA(Low Rank Adaptation)에서는 제2 디지털 미디어 컨텐츠 모델의 파라메타 자체는 고정(freeze)시키면서 파인-튜닝의 학습 대상으로부터 제외시키되, 제2 디지털 미디어 컨텐츠 모델의 파라메타의 조정분에 대해서는 저차원 행렬 분해(low-rank decomposition)를 통하여 파라메타의 조정분의 차원을 당초의 제2 디지털 미디어

컨텐츠 모델의 파라메타의 차원 보다 낮은 저차원 행렬의 행렬 곱의 형태로 표현하여 전체 학습 대상의 파라메타의 개수를 획기적으로 줄일 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 제2 디지털 미디어 컨텐츠 모델에서 2560x2560(dxk)의 차원을 갖는 파라메타 행렬에 대한 저차원 행렬 분해(low rank decomposition)로부터 저차원 행렬인 어뎁터 A의 2560x16(dxr, r=rank)의 차원과 어뎁터 B의 16x2560(kxr, r=rank) 차원을 고려하면, 2560x2560=6,553,600의 파라메타의 개수가 2560x16+16x2560=81,920의 파라메타의 개수로 획기적으로 감소할 수 있다.

【0193】예를 들어, 본 발명의 일 실시형태에서, 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에 적용되는 LoRA에서 기존의 제2 디지털 미디어 컨텐츠 모델의 파라메타 세트는 고정(freeze)시키면서, 파인-튜닝을 통하여 변화되는 파라메타 세트의 조정분에 대해 저차원 행렬 분해(low-rank decomposition)를 통하여 행렬 곱의 형태로 표현된 2개의 서로 다른 저차원 행렬(LoRA에서 학습되는 저차원 행렬)을 형성하는 element를 학습 대상으로 하면서, 제2 디지털 미디어 컨텐츠 모델의 파라메타 자체를 학습 대상으로 하는 종전의 파인-튜닝에서와 달리, 제2 디지털 미디어 컨텐츠 모델이 습득한 지식이 보존될 수 있으며(파인-튜닝 이전의 training된 제2 디지털 미디어 컨텐츠 모델의 지식이 보전됨), 예를 들어, 제2 디지털 미디어 컨텐츠 모델의 마지막 FC layer에 파라메타를 파인-튜닝으로 업-데이트(갱신)시키는 경우에, 제2 디지털 미디어 컨텐츠 모델의 마지막 layer의 지식이 사라질 수 있으나, 본 발명의 일 실시형태에서 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에 적용될

수 있는 LoRA에서는 제2 디지털 미디어 컨텐츠 모델의 파라메타 자체를 고정(freeze)시키므로, 종전 제2 디지털 미디어 컨텐츠 모델의 지식을 그대로 보존하면서도 파인-튜닝을 통하여 학습된 저차원 행렬의 파라메타에 따라 파라메타의 튜닝을 구현할 수 있다.

【0194】보다 구체적으로, 상기 LoRA의 파인-튜닝은 이하와 같은 단계로 구성될 수 있다.

【0195】1) 제2 디지털 미디어 컨텐츠 모델의 파라메타 고정(freeze)

【0196】파인-튜닝에서는 제2 디지털 미디어 컨텐츠 모델의 파라메타 세트가 업-데이트 되지 않으며, LoRA(Low Rank Adaptation) adpater(예를 들어, LoRA에서 학습되는 저차원의 행렬)을 제2 디지털 미디어 컨텐츠 모델의 특정 layer에 추가할 수 있으며, 이때, 상기 LoRA 모듈은 제2 디지털 미디어 컨텐츠 모델의 파라메타의 조정분에 대한 저차원 행렬 분해에 해당되는 저차원 행렬의 행렬 곱을 포함할 수 있다.

【0197】2) Input

【0198】LoRA의 파인-튜닝에서는 파인-튜닝을 위한 학습 데이터(원본 데이터+가상 데이터)가 제2 디지털 미디어 컨텐츠 모델과 LoRA adpater(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)에 함께 입력되며, 제2 디지털 미디어 컨텐츠 모델의 파라메타가 학습되지 않을 수 있다.

【0199】3) Adapter Train

【0200】adapter A(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)를 통과한 학습 데이터는 adapter B(LoRA에서 학습되는 저차원 행렬, LoRA 행렬)를 통과하면서 BA의 행렬 곱이 생성될 수 있으며, 예를 들어, 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬의 차원인 (dxk)의 차원을 갖는 BA 행렬 곱(ΔW=BA)이 생성될 수 있고, 이때, adapter A 및 adapter B의 차원 내지는 파라메타의 개수는 r(rank)로부터 결정될 수 있으며, 각각의 adapter A는 (rxk) 차원과 adapter B는 (dxr) 차원을 가질 수 있다. 이때 adapter A 및 adapter B의 차원 내지는 파라메타의 개수를 결정하는 r(rank)는 제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬의 차원을 형성하는 d 또는 k 보다 작은 값으로 설정될 수 있다(r << d, k).

【0201】4) 최종 출력 산출

【0202】제2 디지털 미디어 컨텐츠 모델의 파라메타 행렬 W를 통과한 출력과 LoRA adapter를 통과한 출력이 합산되면서 최종 출력을 산출할 수 있으며, 최종 출력으로서 W x input + B x A x input의 최종 출력이 산출될 수 있다.

【0203】5) Backpropagation

【0204】4) 단계에서 산출된 최종 출력과 파인-튜닝을 위한 학습 데이터의 세트의 타겟 레이블로부터 Backpropagation을 통하여 adapter A와 adapter B의 파라메타를 업-데이트 시킬 수 있다.

【0205】앞서 설명된 바와 같은 1)~4) 단계는 파인-튜닝 이후의 추론 단계에서도 동일하게 진행될 수 있으며, 상기와 같은 1)~4) 단계를 통하여 최종 출력으로

부터 추론 단계에서의 예측이 이루어질 수 있다.

【0206】본 발명의 다양한 실시형태에서 제2 디지털 미디어 컨텐츠 모델의 파인-튜닝에서는 LoRA(Low Rank Adaptation) 외에 QLoRA(Quantized Low Rank Adaptation)가 적용될 수도 있다. 예를 들어, 상기 QLoRA는 제2 디지털 미디어 컨텐츠 모델의 파라메타를 양자화된 4비트로 로딩하여 메모리의 할당량과 프로세서(GPU)의 연산량을 줄이면서 효율성을 더 높인 LoRA의 개선된 버전으로, 예를 들어, 상기 LoRA의 파인-튜닝에서는 제2 디지털 미디어 컨텐츠 모델의 파라메타가 고정될 수 있으며, QLoRA의 파인-튜닝에서는 제2 디지털 미디어 컨텐츠 모델의 파라메타가 4비트로 고정될 수 있고, 4비트로 로딩된 모델이 파인-튜닝에 적용될 수 있다. 다만 상기 QLoRA의 파인-튜닝에서 제2 디지털 미디어 컨텐츠 모델의 파라메타가 4비트로 고정된다고 하더라도 이후의 추론 과정에서는 8비트로 로딩된 모델이 추론 과정에서 적용될 수 있다.

【0207】상기 QLoRA에서는 4비트 양자화와 double quantization이 적용될 수 있다. 즉, 상기 QLoRA에서는 제2 디지털 미디어 컨텐츠 모델의 파라메타를 4비트 NormalFloat(NF4)라는 새로운 데이터 타입으로 양자화 시킬 수 있으며, 인공지능 모델의 파라메타는 32비트 부동 소수점(float 32) 또는 16비트 부동 소수점(float 16/bfloat 16)으로 표현될 수 있으나, 상기 QLoRA의 파인-튜닝에서는 4비트 양자화를 통하여 메모리의 할당량을 획기적으로 줄일 수 있다. 또한, 상기 QLoRA의 파인-튜닝에서는 양자화 상수(quantization constant)에도 추가적인 양자화를 적용하는 double quantization을 적용할 수 있으며, 양자화 상수는 양자화 과정에서 사용되

는 값으로 메모리의 할당이 요구되므로, 상기와 같은 double quantization을 통하여 메모리의 할당량을 줄일 수 있다.

【0208】 본 발명의 다양한 실시형태에서, 상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서는, 앞서 설명된 바와 같은 LoRA 또는 QLoRA 외에, Mixed precision(FP32와 FP16 혼용), xFormers 메모리 최적화, 그래디언트 누적(Gradient Accumulation, 하나의 mini-batch 마다 그래디언트를 산출하기 보다는 다수의 mini-batch에 걸쳐서 그래디언트를 누적하여 메모리 사용량을 줄이면서 단일 mini-batch에 대한 그래디언트 예측 보다 더 정확한 예측을 제공함), 학습률 스케줄러(학습률을 조절하여 경사 하강법의 업 데이터 속도를 조절함, 학습률 또는 그래디언트의 크기 norm가 큰 경우 발산 방지, 학습률이 작은 경우 수렴 속도 내지는 학습 속도의 지연 방지), 그래디언트 클립핑(Gradient Clipping, 그래디언트의 크기 norm을 제한하여 그래디언트 방향은 유지하면서 발산 방지) 등을 도입하여 안정적이면서 효과적으로 파인-튜닝이 이루어지도록 할 수 있다.

【0209】 도 24a 내지 도 24c에는 본 발명의 일 실시형태에 따른 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성의 일 예시로서, 서로 다른 원본 데이터(원본 데이터로서의 이미지)를 예시적으로 보여주는 도면들이 도시되어 있다.

【0210】 도 25a 내지 도 25i에는 본 발명의 일 실시형태에 따른 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성의 일 예시로서, 도 24a에 예시된 원본 데이터(원본 데이터로서 이미지)에 부여된 서로 다른 변형 조건으로부터 생성된 서로 다른 가상 데이터(가상 데이터로서 이미지)를 예시적으로 보여주는 도면들이 도시

되어 있으며, 보다 구체적으로, 도 25a 내지 도 25i에 예시된 각각의 서로 다른 가상 데이터(가상 데이터로서 이미지)는 서로 다른 변형 조건으로, 1) brightness_down, 2) brightness_up, 3) gray scale, 4) horizontal_flip, 5) rotation_+45 degree, 6) rotation_-45 degree, 7) saturation_down, 8) saturation_up, 9) vertical_flip의 변형 조건으로부터 생성된 서로 다른 가상 데이터를 보여주는 도면들이 도시되어 있다.

【0211】 도 26a 내지 도 26c에는 각각 도 25a 내지 도 25c에 예시된 가상 데이터(가상 데이터로서 이미지)를 합성 데이터(합성 데이터로서 이미지)의 이미지 생성을 가이드 하기 위한 이미지 conditioning으로 하고, 합성 데이터(합성 데이터로서 이미지)의 이미지 생성을 가이드 하기 위한 텍스트 conditioning을 주입하여 생성된 합성 데이터를 예시적으로 보여주는 도면들로서, 하기와 같은 이미지 conditioning 및 텍스트 conditioning을 디퓨전 모델(Diffusion Model, 또는 레이턴트 디퓨전 모델, LDM, Latent Diffusion Model, 또는 스테이블 디퓨전 모델, stable Diffusion Model, 예를 들어, 도 7에 도시된 제1 디지털 미디어 컨텐츠 생성 모델)에 주입하여 생성된 서로 다른 합성 데이터(합성 데이터로서 이미지)를 예시적으로 보여주는 도면이 도시되어 있다.

【0212】 도 26a:

【0213】 (이미지 conditioning) 도 24a의 가상 데이터

【0214】(텍스트 conditioning) "A mesmerizing abstract constellation floating in deep cosmic blues and purples, stars and swirling nebulae forming intricate patterns across a vast universe"

【0215】도 26b:

【0216】(이미지 conditioning) 도 24b의 가상 데이터

【0217】(텍스트 conditioning) "A vibrant abstract depiction of constellations interwoven with shimmering stardust, set against a backdrop of swirling galactic colors and cosmic haze"

【0218】도 26c:

【0219】(이미지 conditioning) 도 24c의 가상 데이터

【0220】(텍스트 conditioning) "An ethereal tapestry of stars and geometric constellation shapes, floating in a radiant universe painted with deep blues, violets, and bursts of light"

【0221】본 발명은 첨부된 도면에 도시된 실시예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 수 있을 것이다.

【부호의 설명】

【0222】Flip, Gray scale, Saturation, Brightness, Rotation: Img2Img의 transformation

Noising process: 노이징 프로세스

Denosing process: 디노이징 프로세스

Contrastive learning: contrastive language image pre-training, CLIP

AdaIN: Adaptive Instant Normalization

LoRA: Low Rank Adaptation

Quantization: 양자화

【청구범위】

【청구항 1】

원본 데이터로부터 원본 데이터의 스타일이 전이된 가상 데이터를 생성하기 위한 것으로, 스타일 전이된 이미지를 생성하도록 training된 제1 디지털 미디어 컨텐츠 생성 모델; 및

상기 원본 데이터와 상기 제1 디지털 미디어 컨텐츠 생성 모델로부터 생성된 가상 데이터를 포함하는 동일 유사 스타일로 표현된 학습 데이터로부터, 스타일 전이된 이미지를 생성하도록 training된 모델을 대상으로 파인-튜닝 시키기 위한 제2 디지털 미디어 컨텐츠 생성 모델;을 포함하고,

상기 제2 디지털 미디어 컨텐츠 생성 모델은 학습 또는 파인-튜닝을 위하여 loss function의 산출을 위한 loss computing 네트워크를 포함하여 loss computing 프로세스를 구현하되,

상기 제1 디지털 미디어 컨텐츠 생성 모델은 상기 loss computing 네트워크를 포함하지 않거나 또는 loss computing 프로세스를 구현하지 않는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 2】

제1항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은, 상기 제1 디지털 미디어 컨텐츠 생성 모델의 네트워크를 공유하는 것을 특징으로 하는, 디지털 미디어 컨텐츠

생성 시스템.

【청구항 3】

제2항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은,

상기 제1 디지털 미디어 컨텐츠 생성 모델을 포함하고,

상기 loss computing 네트워크를 더 포함하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 4】

제1항에 있어서,

상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터(y)로서 원본 데이터 또는 가상 데이터와, 생성 대상 이미지에 표현될 컨텐츠 정보를 포함하는 컨텐츠 데이터(x)를 서로 다른 입력으로 하여,

컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 instant normalization 된 컨텐츠 데이터(t)로부터 Generator G(g)를 통하여 이미지 또는 feature를 생성하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 5】

제4항에 있어서,

상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

상기 스타일 데이터(y)로부터 추출된 feature statistics로부터 추출된 스케일링(scaling) 파라메타 및 바이어스(bias) 파라메타를 적용한 AdaIN(Adaptive Instant Normalization)을 통하여 instant normalization된 컨텐츠 데이터(t)를 생성하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 6】

제5항에 있어서,

상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 이미지 또는 feature를 출력하기 위한 Generator G(g)로서, convolution layer 또는 업-스케일링을 위한 Upsample 네트워크를 포함하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 7】

제6항에 있어서,

상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

저해상도의 feature 또는 feature map으로부터 고해상도의 이미지를 향하여 progressive 하게 업-스케일링으로 전개되는 멀티-스케일을 형성하는 각각의 스케일에서, 상기 Generator G(g)로부터 이미지 또는 feature의 생성이 구현되는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 8】

제7항에 있어서,

상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

레이턴트 표현(z)을 입력으로 하여 레이턴트 표현(z)의 서로 다른 차원 사이의 disentanglement를 위한 다수의 FC(Fully Connected) layer의 적층과, 상기 다수의 FC layer의 적층으로부터의 출력과 상기 원본 데이터 또는 가상 데이터를 서로 다른 입력으로 하여, 상기 Generator G(g)를 포함하는 Synthesis network로 주입되는 스타일 정보를 추출하기 위한 intermediate 레이턴트 표현(w)을 생성하기 위한 Mapping network를 포함하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 9】

제5항에 있어서,

상기 제1 디지털 미디어 컨텐츠 생성 모델 또는 제2 디지털 미디어 컨텐츠 생성 모델은,

상기 컨텐츠 데이터(x) 및 스타일 데이터(y)로부터 feature를 추출하기 위한 인코더(Encoder); 및

상기 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t)를 입력으로 하여 이미지(g(t)) 또는 feature(g(t))를 출력하기 위한 Generator G(g)로서, AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t)로부터 차원 확장을 통하여 이미지를 생성하기 위한 디코더(Decoder);를 포함하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 10】

제1항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은,

생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터(y)로서 원본 데이터와 생성 대상 이미지에 표현될 컨텐츠 정보를 포함하는 컨텐츠 데이터(x)를 서로 다른 입력으로 하여,

컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)을 통하여 instant normalization된 컨텐츠 데이터(t)로부터 Generator G(g)에서 생성된 이미지(g(t)) 또는 feature(g(t))를 입력으로 하는 loss computing 네트워크의 출력(f(g(t))과, AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t) 사이의 contents loss(Lc)와, 상기 loss computing 네트워크의 출력(f(g(t))과 스타일 데이터(y) 사이의 style loss(Ls)를 취합한 loss function을 산출하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 11】

제10항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은,

상기 loss computing 네트워크의 출력(f(g(t))과 AdaIN(Adaptive instant Normalization)으로 instant normalization된 컨텐츠 데이터(t) 사이의 contents loss(Lc)로서, 상기 loss computing 네트워크의 출력(f(g(t))과 AdaIN으로 instant normalization된 컨텐츠 데이터(t)의 feature 또는 feature map 사이에서 이하와 같은 픽셀 단위의 거리(L2 norm)에 기반한 contents loss(Lc)를 산출하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 12】

제10항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은,

상기 loss computing 네트워크의 출력(Φ(g(t))로부터 추출된 feature statistics로서 mean(평균) μ와 standard deviation(표준편차) σ와, 스타일 데이터(y)로부터 추출된 feature statistics로서 mean(평균) μ와 standard deviation(표준편차) σ 사이의 거리(L2 norm)에 기반한 이하와 같은 style loss(Ls)를 산출하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 13】

제10항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은, 상기 contents loss(Lc) 및 style loss(Ls)가 스케일 팩터(λ)를 적용하여 합산된 이하와 같은 loss function을 산출하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 14】

제10항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은, loss computing 네트워크로부

터 산출된 loss function을 최소화시키도록 학습되는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 15】

제10항에 있어서,

상기 loss function을 산출하기 위한 loss computing 네트워크는, Generator G(g)의 역-프로세스를 구현하면서,

Generator G(g)로부터 생성된 이미지(g(t)) 또는 feature(g(t))에 대해 역-프로세스를 적용하여, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN을 통하여 instant normalization된 컨텐츠 데이터(t)로 복원되는지에 관한 contents loss(Lc); 및

Generator G(g)로부터 생성된 이미지(g(t)) 또는 feature(g(t))에 대해 역-프로세스를 적용하여, 스타일 데이터(y) 또는 스타일 데이터(y)로부터 추출된 feature statistics를 추종하는지에 관한 style loss(Ls)를 산출하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 16】

제15항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델의 파인-튜닝에서, 상기 Generator G(g)는 학습 대상에 해당되되, 상기 loss computing 네트워크는 학습 대상에 해당되지 않는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 17】

제15항에 있어서,

상기 Generator G(g)는, 컨텐츠 데이터(x) 또는 컨텐츠 데이터(x)로부터 추출된 feature에 대해, 스타일 데이터(y)로부터 추출된 feature statistics를 적용한 AdaIN(Adaptive instant Normalization)을 통하여 instant normalization된 컨텐츠 데이터(t)로부터 차원 확장을 통하여 이미지(g(t)) 또는 feature(g(t))를 생성하기 위한 디코더(Decoder)를 포함하고,

상기 loss computing 네트워크는, 상기 디코더(Decoder)의 역-프로세스로서 디코더(Decoder)로부터 생성된 이미지(g(t)) 또는 feature(g(t))로부터 차원 축소된 feature를 추출하기 위한 인코더(Encoder)를 포함하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 18】

제17항에 있어서,

상기 제2 디지털 미디어 컨텐츠 생성 모델은,

상기 컨텐츠 데이터 및 스타일 데이터로부터 feature를 추출하기 위한 제1 인코더를 더 포함하고,

상기 loss computing 네트워크는 상기 제1 인코더와 동일한 네트워크로 구현되는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 19】

제18항에 있어서,

상기 제1 인코더 및 loss computing 네트워크는 CNN(Convolution Neural Network) 계열의 VGG net 기반의 Autoencoder로서 사전에 학습된 VGG net의 convolution layer를 포함하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 20】

제1항에 있어서,

상기 제1 디지털 미디어 컨텐츠 모델 또는 제2 디지털 미디어 컨텐츠 모델은,

원본 이미지로부터 시간 스텝을 전진시키면서 노이즈 스케줄(noise schedule)로부터 정의된 노이즈를 추가하여 점진적으로 원본 이미지의 패턴을 붕괴시키는 노이징 프로세스(noising process)의 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복원되도록 노이지한 이미지로부터 상대적으로 원본 이미지에 가까운 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현하면서 이미지를 생성하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 21】

제20항에 있어서,

상기 제1 디지털 미디어 컨텐츠 모델과 제2 디지털 미디어 컨텐츠 모델은,

노이즈를 추가하면서 원본 이미지를 붕괴시키는 노이징 프로세스(noising process)의 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복원되도록 노이지한 이미지로부터 상대적으로 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현하기 위한 네트워크를 공유하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 22】

제21항에 있어서,

상기 제1 디지털 미디어 컨텐츠 모델은,

이미지 생성을 위한 denoising process를 구현하기 위한 네트워크를 포함하여 denoising process는 수행하되,

학습 또는 파인-튜닝을 위한 forward process로서 noising process를 구현하기 위한 네트워크는 포함하지 않거나 또는 학습 또는 파인-튜닝을 위한 forward process로서 noising process는 수행하지 않는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 23】

제21항에 있어서,

상기 제2 디지털 미디어 컨텐츠 모델은,

이미지 생성을 위한 denoising process를 구현하기 위한 네트워크 및 학습

또는 파인-튜닝을 위한 forward process로서 noising process를 구현하기 위한 네트워크를 모두 포함하여, 이미지 생성을 위한 denoising process와 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 모두 수행하면서,

상기 denoising process에서 현재의 시간 스텝에서의 이미지(Xt)로부터 덜 노이지한 이전 시간 스텝(Xt-1)에서 추가된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균을 예측하고, noising process에서 산출된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균과의 오차를 최소화 시키도록 상기 denoising process를 학습하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 24】

제20항에 있어서,

상기 제1 디지털 컨텐츠 생성 모델은,

이미지 생성 조건으로 이미지 conditioning 또는 텍스트 conditioning의 주입을 위한 인코더를 포함하는 것을 특징으로 하는, 디지털 컨텐츠 생성 시스템.

【청구항 25】

제24항에 있어서,

상기 인코더는,

상기 이미지 생성 조건으로, 상기 원본 데이터를 이미지 conditioning으로 주입하기 위한 이미지 인코더; 및

상기 이미지 생성 조건으로, 상기 원본 데이터의 변형 조건 또는 가상 데이

터로 표현될 컨텐츠 정보를 텍스트 conditioning으로 주입하기 위한 텍스트 인코더;를 포함하는 것을 특징으로 하는, 디지털 컨텐츠 생성 시스템.

【청구항 26】

제25항에 있어서,

상기 이미지 인코더 및 텍스트 인코더는 각각,

멀티-모달의 CLIP(contrastive language image pre-training)을 통하여 이미지-텍스트의 멀티-모달(multi modal)의 임베딩 공간을 학습하여,

이미지 conditioning으로 주입된 원본 데이터를 클립 이미지 임베딩(CLIP image embedding)으로 인코딩하고,

텍스트 conditioning으로 주입된 원본 데이터의 변형 조건 또는 가상 데이터로 표현될 컨텐츠 정보를 클립 텍스트 임베딩(CLIP text embedding)으로 인코딩하는 것을 특징으로 하는, 디지털 컨텐츠 생성 시스템.

【청구항 27】

제24항에 있어서,

상기 제1 디지털 미디어 컨텐츠 모델은,

노이즈를 추가하면서 원본 이미지를 붕괴시키는 노이징 프로세스(noising process)의 역-프로세스(reverse process)로서 원본 이미지의 패턴이 복원되도록 노이지한 이미지로부터 상대적으로 덜 노이지한 이미지를 생성하는 디노이징 프로세스(denoising process)를 구현하면서,

상기 이미지 생성 조건을 주입하기 위한 인코더를 포함하여, 이미지 conditioning 또는 텍스트 conditioning을 이미지 임베딩 또는 텍스트 임베딩으로 인코딩하여 생성 조건으로 주입하기 위한 임베딩 또는 인코딩을 구현하는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【청구항 28】

제20항에 있어서,

상기 제2 디지털 미디어 컨텐츠 모델은,

이미지 생성을 위한 denoising process를 구현하기 위한 네트워크 및 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 구현하기 위한 네트워크를 모두 포함하여, 이미지 생성을 위한 denoising process와 학습 또는 파인-튜닝을 위한 forward process로서 noising process를 모두 수행하면서,

상기 denoising process에서 현재의 시간 스텝에서의 이미지(Xt)로부터 덜 노이지한 이전 시간 스텝(Xt-1)에서 추가된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균을 예측하고, noising process에서 산출된 노이즈 또는 덜 노이지한 이전 시간 스텝(Xt-1)의 평균과의 오차를 최소화 시키도록 상기 denoising process를 학습하되,

상기 이미지 생성 조건을 주입하기 위한 인코더를 포함하지 않거나 또는 이미지 conditioning 또는 텍스트 conditioning을 이미지 임베딩 또는 텍스트 임베딩으로 인코딩하여 생성 조건으로 주입하기 위한 임베딩 또는 인코딩을 구현하지 않

는 것을 특징으로 하는, 디지털 미디어 컨텐츠 생성 시스템.

【요약서】

【요약】

본 발명에서는 데이터 증강에 기반한 디지털 미디어 컨텐츠 생성 시스템이 개시된다. 본 발명에 의하면, 원본 데이터의 희귀성에도 불구하고 생성형 AI 모델의 파인-튜닝을 위한 충분한 학습 데이터를 확보하고 확보된 파인-튜닝을 위한 학습 데이터로부터 과적합(over-fitting)이 없이 일반화 능력이 향상되면서, 생성 대상 이미지에 표현될 주제 또는 객체 등에 관한 컨텐츠 정보를 포함하는 이미지 또는 텍스트의 컨텐츠 데이터와, 생성 대상 이미지로 전이될 스타일 정보를 포함하는 스타일 데이터를 입력으로 하여, 컨텐츠 데이터에 표현된 생성 대상 이미지의 주제 또는 객체 등에 관한 컨텐츠가 스타일 데이터로부터 추출된 스타일로 표현되는 합성 데이터를 생성할 수 있는, 생성형 AI 네트워크가 구현될 수 있다.

【대표도】

도 1

【도면】

【도 1】

【도 2】

【도 3a】

【도 3b】

【도 4】

【도 5】

【도 6】

【도 7】

【도 8】

【도 9】

【도 10】

【도 11】

【도 12】

【도 13】

【도 14】

【도 15】

【도 16】

【도 17】

【도 18】

【도 19a】

【도 19b】

【도 20】

【도 21】

【도 22】

【도 23】

【도 24a】

【도 24b】

【도 24c】

【도 25a】

【도 25b】

【도 25c】

【도 25d】

【도 25e】

【도 25f】

【도 25g】

【도 25h】

【도 25i】

【도 26a】

【도 26b】

【도 26c】

프로젝트 개요

본 시스템은 이미지 큐레이션 텍스트를 생성하는 모델을 활용하여 디지털 미디어 콘텐츠의 큐레이션 텍스트를 자동으로 생성하는 시스템입니다. CLIP과 BLIP-2 모델 기반으로 이미지 특징을 분석하고, 분류 및 키워드 추출, 사용자 메타 정보를 반영해 큐레이터 스타일의 설명 텍스트를 자동 생성합니다.

주요 기능 및 역할

- 구글 드라이브 마운트: 이미지 데이터가 저장된 경로 마운트
- 모델 로딩: CLIP(ViT-L/14)과 BLIP-2(OPT-2.7B) 모델 로드 및 GPU 최적화
- 마스터 택소노미 구조: 박물관/예술 작품 분류를 위한 상세 분류체계(SCENE_CLASSES 등) 및 상호배타 규칙 관리
- 이미지 라벨링: CLIP 모델을 사용한 이미지 특징 추출과 택소노미 기반 라벨 선정
- BLIP-2 및 JSON 기반 큐레이션 텍스트 생성: BLIP-2 모델에 사용자 메타 정보(재료, 기법 등)를 포함하여 큐레이터 스타일의 자연어 및 JSON 텍스트 생성
- 배타 규칙 적용 및 동의어 정규화: 라벨 간 중복 및 상호 배타성 관리, 동의어 자동 변환
- 큐레이터 스타일 문장 렌더링: 시각적 특징을 바탕으로 자연스럽고 통일감 있는 설명 텍스트 생성 및 해시태그 자동 부여
- 이미지 리스트 관리 및 자동 출력: 이미지 디렉토리 내 이미지들을 정렬 후 처리 및 출력
- 안전망 필터링: 특정 키워드(예: url, auction 등) 제거로 안전한 설명문 생성

사용 기술

- Python 3.x
- PyTorch, transformers (HuggingFace)
- CLIP, BLIP-2 pretrained models
- BitsAndBytes (양자화 8bit 모델 로드)
- PIL, numpy
- Google Colab, Google Drive 연동

시스템 구성도 (요약)

1. 이미지 데이터 로드 및 전처리
2. CLIP 기반 시각적 라벨 평가
3. BLIP-2 기반 상세 큐레이션 텍스트 생성

4. 사용자 메타 기입 및 배타 규칙 적용
5. 최종 큐레이터 스타일 자연어 및 태그 생성
6. 결과 출력 및 필터링

담당 역할

- ART/Museum 큐레이션을 위한 이미지 기반 자동 텍스트 생성 시스템 설계 및 구현
- HuggingFace transformers 라이브러리 기반 CLIP, BLIP-2 대규모 모델 효과적 로드 및 최적화
- 복잡한 택소노미 구조 및 상호배타 규칙을 파이썬 코드로 효율적으로 구현
- 이미지 특징 추출 및 라벨링, 텍스트 생성 파이프라인 구축
- 사용자 메타 데이터를 자동으로 반영하는 큐레이터 스타일 자연어 생성 기능 개발
- Google Colab 기반 환경 구축, 구글 드라이브 연동 및 데이터 자동 처리
- NLP 및 컴퓨터 비전 기술을 융합한 멀티모달 연구 및 구현 경험

구현 성과

- 복합 택소노미와 배타 규칙 관리로 정확하고 일관성 있는 큐레이션 텍스트 생성
- BLIP-2 모델 양자화(8bit) 활용으로 GPU 메모리 효율화 및 성능 최적화
- 자동 해시태그 및 큐레이터 톤 자연어 렌더링으로 UX 향상
- 수백 장 이미지에 대한 자동 처리 및 설명문 생성 자동화로 연구 효율성 극대화
- 안전망 필터링 적용으로 부적절한 텍스트 배제

주요 사용 기술 및 도구

- Python, PyTorch, Transformers, BitsAndBytes
- HuggingFace pretrained models: CLIP, BLIP-2
- Google Colab, Google Drive
- 자연어 처리 및 컴퓨터 비전 융합 기술
- 배타 규칙 및 텍스트 렌더링 로직

【서지사항】

【서류명】 특허출원서

【참조번호】 PN159878KR

【출원구분】 특허출원

【출원인】

【성명】 두지언

【특허고객번호】 4-2023-000001-9

【대리인】

【명칭】 리앤목 특허법인

【대리인번호】 9-2005-100002-8

【지정된변리사】 이영필,이해영,민관호

【발명의 국문명칭】 큐레이션 생성 시스템

【발명의 영문명칭】 A system for generating curation

【발명자】

【성명】 두지언

【특허고객번호】 4-2023-000001-9

【발명자】

【성명】 이정민

【성명의 영문표기】 LEE, Jung Min

【국적】 KR

【주민등록번호】 900723-0XXXXXX

【우편번호】 35224

【주소】 대전광역시 서구 계룡로319번길 13 (월평동)

【거주국】 KR

【발명자】

【성명】 이세현

【성명의 영문표기】 LEE, Se Hyeon

【국적】 KR

【주민등록번호】 980402-0XXXXXX

【우편번호】 50917

【주소】 경상남도 김해시 가락로125번길 40, 301호 (봉황동)

【거주국】 KR

【출원언어】 국어

【심사청구】 청구

【취지】 위와 같이 특허청장에게 제출합니다.

대리인 리앤목 특허법인 (서명 또는 인)

【수수료】

【출원료】 0 면 46,000 원

【가산출원료】 117 면 0 원

【우선권주장료】 0 건 0 원

【심사청구료】 27 항 1,543,000 원

【합계】 1,589,000원

【감면사유】 개인(70%감면)[1]

【감면후 수수료】 476,700 원

【수수료 자동납부번호】 3010358854671

【첨부서류】 1.기타첨부서류[위임장]_1통

【발명의 설명】

【발명의 명칭】

큐레이션 생성 시스템{A system for generating curation}

【기술분야】

【0001】 본 발명은 큐레이션 생성 시스템에 관한 것이다.

【발명의 배경이 되는 기술】

【0002】 AI 기술의 발전에 따라 text-to-image, image-to-text와 같은 멀티 모달의 AI 모델로부터 서로 다른 이종 모달의 텍스트, 이미지 등의 미디어를 생성할 수 있는 생성형 AI 모델(generative AI model)이 개발되고 있으며, 이러한 생성형 AI 모델은 의료, 금융, 게이밍, 마케팅, 패션과 같은 다양한 분야에서의 활용이 모색되고 있는 실정이며, 최근에는 전시 공간에 대한 추천과 전시 대상 작품에 대한 설명이나 서술을 제공하여 전시 대상 작품에 대한 이해를 돕고 가치를 부여하도록 또는 이를 통하여 전시 대상 작품을 소비할 수 있도록 돕기 위한 목적으로, 텍스트 시퀀스 형태로 제공되는 전시 대상 작품에 대한 설명이나 서술을 제공하기 위한 AI 큐레이터 개발이 요구되고 있다.

【발명의 내용】

【해결하고자 하는 과제】

【0003】 본 발명의 일 실시형태는 전시 대상 작품을 형성하는 이미지 및 전시 대상 작품의 이미지에 부수되는 메타 데이터로서, 전시 대상 작품을 기획한 작

가의 작업 의도, 작가의 창작 의도, 또는 작가의 철학이 포함된 작가 노트에 관한 메타 데이터 및/또는 전시 대상 작품에 대한 전문가의 작품 평, 작가 또는 전문가가 명명한 작품 명, 전시 대상 작품의 사이즈, 또는 전시 대상 작품을 형성하는 재료에 관한 매체 정보를 포함하는 작품 캡션 정보에 관한 메타 데이터로부터 전시 대상 작품의 서로 다른 특징들을 추출하여 텍스트 형태의 설명 또는 서술을 포함하는 협의의 큐레이션을 생성할 수 있으며, 또한 이와 같이 생성된 협의의 큐레이션을 입력으로 하여 전시 대상 작품의 전시 공간에 관한 추천을 포함하는 광의의 큐레이션을 생성할 수 있는 큐레이션 생성 시스템을 포함한다.

【과제의 해결 수단】

【0004】 상기와 같은 과제 및 그 밖의 과제를 해결하기 위하여, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은,

【0005】 전시 대상 작품을 형성하는 이미지 데이터 및 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터로부터 전시 대상 작품의 서로 다른 특징들을 추출하여 텍스트 형태의 설명 또는 서술을 포함하는 큐레이션을 생성할 수 있다.

【0006】 예를 들어, 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터로서 텍스트 데이터는,

【0007】 전시 대상 작품을 기획한 작가의 작업 의도, 작가의 창작 의도, 또는 작가의 철학이 포함된 작가 노트에 관한 텍스트 데이터; 또는

【0008】 전시 대상 작품에 대한 전문가의 작품 평, 작가 또는 전문가가 명명한 작품 명, 전시 대상 작품의 사이즈, 또는 전시 대상 작품을 형성하는 재료에 관한 매체 정보를 포함하는 작품 캡션 정보에 관한 텍스트 데이터를 포함할 수 있다.

【0009】 예를 들어, 상기 큐레이션 생성 시스템은,

【0010】 전시 대상 작품의 서로 다른 특징들을 추출하기 위한 다수의 분석 모델; 및

【0011】 상기 다수의 분석 모델로부터 추출된 서로 다른 특징들을 취합하여 상기 큐레이션을 생성하기 위한 신경망 네트워크를 포함할 수 있다.

【0012】 예를 들어, 상기 다수의 분석 모델은, 상기 전시 대상 작품을 형성하는 이미지 상에 표현된 서로 다른 특징들을 추출하기 위한 것으로,

【0013】 색감 또는 컬러 톤에 관한 특징을 추출하기 위한 색감 분석 모델;

【0014】 감성에 관한 특징을 추출하기 위한 감성 분석 모델;

【0015】 스타일에 관한 특징을 추출하기 위한 스타일 분석 모델;

【0016】 전시 대상 작품 상에 표현된 객체(object)에 관한 특징을 추출하기 위한 객체 분석 모델; 및

【0017】 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 주제, 테마, 분위기 또는 감성을 포괄하는 의도에 관한 특징을 추출하기 위한 의도 분석 모델:을 포함할 수 있다.

【0018】예를 들어, 상기 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델, 객체 분석 모델 및 의도 분석 모델은, 전시 대상 작품을 형성하는 이미지 데이터 또는 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여, 동일한 텍스트 임베딩 공간 상으로 매핑되는 텍스트 임베딩 표현을 출력할 수 있다.

【0019】예를 들어, 상기 다수의 분석 모델 각각에 대한 입출력 관계에 관하여,

【0020】상기 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델 및 객체 분석 모델은, 전시 대상 작품의 이미지 데이터를 입력으로 하여 서로 다른 특징에 관한 텍스트 임베딩 표현을 출력하면서 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하고,

【0021】상기 의도 분석 모델은, 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여 입력된 메타 데이터의 요약 생성을 통하여 텍스트 임베딩 표현을 출력하면서 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현할 수 있다.

【0022】예를 들어, 상기 감성 분석 모델은, 상기 감성 분석 모델에 대한 입출력 관계에 관하여,

【0023】전시 대상 작품의 이미지 데이터를 입력으로 하여 감성 분석의 예측에 관한 텍스트 임베딩 표현을 출력하면서 이미지-투-텍스트의 멀티-모달(multi-

modal)을 구현하는 이미지 기반의 감성 분석 모델을 포함할 수 있다.

【0024】 예를 들어, 상기 감성 분석 모델은, 상기 감성 분석 모델에 대한 입출력 관계에 관하여,

【0025】 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여 입력된 메타 데이터로부터 감성 분석의 예측에 관한 텍스트 임베딩 표현을 출력하면서 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델을 더 포함할 수 있다.

【0026】 예를 들어, 상기 감성 분석 모델은, 상기 이미지 기반의 감성 분석 모델의 예측 결과와, 상기 텍스트 기반의 감성 분석 모델의 예측 결과를 취합하여 전시 대상 작품으로부터 추출된 감성에 관한 특징으로 출력하기 위한 신경망 네트워크를 더 포함할 수 있다.

【0027】 예를 들어, 상기 의도 분석 모델은, 상기 메타 데이터로서 텍스트 시퀀스, 단어 집합 또는 문장 집합으로부터 입력된 메타 데이터를 함축한 요약 생성을 구현할 수 있다.

【0028】 예를 들어, 상기 의도 분석 모델은, 적어도 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 감성에 관한 특징을 추출하고,

【0029】 상기 감성 분석 모델은,

【0030】 전시 대상 작품의 이미지 데이터를 입력으로 하는 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하는 이미지 기반의 감성 분석 모델은 포함하되,

【0031】 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델은 포함하지 않을 수 있다.

【0032】 예를 들어, 상기 의도 분석 모델은, 적어도 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 감성에 관한 특징을 추출하고,

【0033】 상기 감성 분석 모델은,

【0034】 전시 대상 작품의 이미지 데이터를 입력으로 하는 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하는 이미지 기반의 감성 분석 모델과 함께,

【0035】 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델을 포함하되,

【0036】 상기 의도 분석 모델과 텍스트 기반의 감성 분석 모델은, 입력된 메타 데이터를 함축한 요약 생성을 위한 네트워크를 공유할 수 있다.

【0037】 예를 들어, 상기 텍스트 기반의 감성 분석 모델은 입력된 메타 데이터를 함축한 요약 생성에 대해, 추출 대상이 되는 감성 분석의 특징과 관련된 임베딩 표현에 상대적으로 높은 가중치를 부여하면서 감성 분석의 특징과 무관한 임베딩 표현에 상대적으로 낮은 가중치를 부여하거나 또는 가중치를 부여하지 않으면서 필터링-아웃(filtering-out)시키도록 학습된 가중치 세트를 포함할 수 있다.

【0038】 예를 들어, 상기 색감 분석 모델은,

【0039】전시 대상 작품을 형성하는 이미지로부터 색감 또는 컬러 톤(color tone) 정보를 표현하기 위한 히스토그램(histogram) 또는 히스토그램 정보에 기반하여 전시 대상 작품을 형성하는 전체 이미지 상에서 표현되는 색감 또는 컬러 톤 정보를 추출할 수 있다.

【0040】예를 들어, 상기 색감 분석 모델은,

【0041】전시 대상 작품의 이미지를 형성하도록 서로에 대해 합성되는 3채널 이미지(R,G,B 3채널 이미지 또는 Y,Cb,Cr 3채널 이미지) 각각에 대해 화소 값 별로 등장하는 화소 개수를 표현하는 제1 내지 제3 히스토그램 또는 제1 내지 제3 히스토그램 정보에 기반하여 전시 대상 작품의 색감 또는 컬러 톤에 관한 특징을 추출할 수 있다.

【0042】예를 들어, 상기 스타일 분석 모델은,

【0043】전시 대상 작품을 형성하는 이미지로부터 추출된 스타일 정보와, 사전에 설정된 다수의 템플릿 이미지 각각으로부터 추출된 스타일 정보 사이의 유사도 분석에 기반하여, 전시 대상 작품의 스타일에 관한 특징을 추출할 수 있다.

【0044】예를 들어, 상기 스타일 분석 모델은,

【0045】전시 대상 작품을 형성하는 이미지와 가장 높은 유사도 스코어가 산출된 템플릿 이미지를 전시 대상 작품과 매칭되는 템플릿 이미지로 하여, 매칭된 템플릿 이미지와 연계된 스타일에 관한 특징을 탐색하여 탐색된 결과에 따라 전시 대상 작품의 스타일에 관한 특징으로 출력할 수 있다.

【0046】예를 들어, 상기 객체 분석 모델은,

【0047】전시 대상 작품의 이미지 상에 등장하는 객체에 대한 인식 또는 분류와 함께, 전시 대상 작품의 이미지 상에 등장하는 다수의 객체들 사이의 상대적인 위치 관계 및 대소 관계를 포함하여 다수의 객체들 사이의 공간 배치를 예측할 수 있다.

【0048】예를 들어, 상기 객체 분석 모델로부터 산출되는 공간 배치의 예측 결과로부터,

【0049】전시 대상 작품의 이미지 상에서 중앙 위치에 인접하게 배치되는 객체일수록, 또는 상대적으로 넓은 영역을 점유하는 객체일수록, 전시 대상 작품의 주체 또는 테마와 인접한 주된 객체로 추론하며,

【0050】전시 대상 작품의 이미지 상에서 중앙 위치로부터 멀리 떨어진 객체일수록, 또는 상대적으로 좁은 영역을 점유하는 객체일수록, 전시 대상 작품의 주체 또는 테마로부터 먼 보조 객체로 추론할 수 있다.

【0051】예를 들어, 본 발명의 다른 실시형태에 따른 큐레이션 생성 시스템은,

【0052】생성 대상인 전시 대상 작품의 큐레이션을 시간 스텝의 전진에 따라 다음에 등장할 표현을 예측하면서 순차적으로 생성되는 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터로 생성하기 위한 Time series based GAN을 포함할 수 있다.

【0053】 예를 들어, 상기 Time series based GAN으로서 Sequence GAN을 형성하는 Generator G와 Discriminator D 중에서 적어도 어느 하나의 네트워크는, RNN(Recurrent Neural Network) 또는 LSTM(Long Short Term memory)의 시퀀스 아키텍처를 포함할 수 있다.

【0054】 예를 들어, 상기 Generator G는 시간 스텝의 전진에 따라 각각의 시간 스텝 마다 random 노이즈 또는 latent vector를 입력으로 하여 각각의 시간 스텝에서 다음에 등장할 표현에 관한 예측으로부터 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터로서 전시 대상 작품의 큐레이션을 순차적으로 생성할 수 있다.

【0055】 예를 들어, 상기 Discriminator D는 Generator G로부터 시간 스텝의 전진에 따라 각각의 시간 스텝 마다 생성된 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터를 입력으로 하여, 각각의 시간 스텝에서의 출력을 산출하되, 각각의 시간 스텝에서의 출력들에 대한 vote를 통하여 최종적인 출력으로서 real과 fake의 판단을 출력할 수 있다.

【0056】 예를 들어, 상기 Generator G는 강화 학습(RL, Reinforcement Learning)의 에이전트(agent)로서,

【0057】 상기 Discriminator D로부터 상대적으로 많은 보상(reward)이나 또는 양(positive)의 보상이 제공되면 향후의 상태 관찰(state observation)로서, 이전 시간 스텝에서 예측된 모든 표현에 관한 관찰을 통하여 동일 유사한 상태에서 동일한 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현으로 동일한 표

현을 출력할 확률을 높이는 방향으로 정책 함수(policy function)를 갱신하며,

【0058】 상기 Discriminator D로부터 상대적으로 적은 보상이나 또는 음(negative)의 보상이 제공되면 향후의 상태 관찰(state observation)로서, 이전 시간 스텝에서 예측된 모든 표현에 관한 관찰을 통하여 동일 유사한 상태에서 동일한 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현으로 동일한 표현을 출력할 확률을 낮추는 방향으로 정책 함수(policy function)를 갱신할 수 있다.

【0059】 예를 들어, 상기 Discriminator D는 시간 스텝의 전진에 따라 매 시간 스텝 마다 상기 Generator G로부터 취해진 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현에 대한 예측에 대해 즉각적인 보상(immediate reward)을 제공하지 않고,

【0060】 하나의 에피소드(Episode)가 종료된 이후로서, 전시 대상 작품의 큐레이션을 형성하는 열 단위 또는 문장 단위의 예측이 종료된 이후에 비로소 상기 Generator G로부터 취해진 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현에 대한 지연된 보상을 제공할 수 있다.

【0061】 예를 들어, 상기 Generator G의 학습에서, 상기 Generator G로부터 취해진 액션으로서, 전시 대상 작품의 큐레이션을 형성할 다음 표현에 관한 현재 시간 스텝에서의 예측 이후로 다음 시간 스텝에서의 예측으로부터 하나의 에피소드(Episode)의 종료까지의 Generator G로부터 취해지는 다음 시간 스텝 이후의 예측은 MC Search Tree(Monte Carlo Search Tree)로부터 예측될 수 있다.

【0062】예를 들어, 상기 Discriminator D는 시간 스텝의 전진에 따라 상기 Generator G로부터 취해진 액션으로서, 전시 대상 작품의 큐레이션을 형성할 다음 표현에 관한 현재 시간 스텝에서의 예측과, 전시 대상 작품의 큐레이션을 형성할 다음 시간 스텝 이후로 하나의 에피소드(Episode)의 종료까지 MC Search Tree(Monte Carlo Search Tree)로부터의 예측을 취합한 에피소드(Episode) 단위로 보상(reward)을 제공할 수 있다.

【발명의 효과】

【0063】본 발명에 의하면, 전시 대상 작품을 형성하는 이미지 및 전시 대상 작품의 이미지에 부수되는 메타 데이터로서, 전시 대상 작품을 기획한 작가의 작업 의도, 작가의 창작 의도, 또는 작가의 철학이 포함된 작가 노트에 관한 메타 데이터 및/또는 전시 대상 작품에 대한 전문가의 작품 평, 작가 또는 전문가가 명명한 작품 명, 전시 대상 작품의 사이즈, 또는 전시 대상 작품을 형성하는 재료에 관한 매체 정보를 포함하는 작품 캡션 정보에 관한 메타 데이터로부터 전시 대상 작품의 서로 다른 특징들을 추출하여 텍스트 형태의 설명 또는 서술을 포함하는 협의의 큐레이션을 생성할 수 있으며, 또한 이와 같이 생성된 협의의 큐레이션을 입력으로 하여 전시 대상 작품의 전시 공간에 관한 추천을 포함하는 광의의 큐레이션을 생성할 수 있는 큐레이션 생성 시스템이 제공된다.

【도면의 간단한 설명】

【0064】도 1a 및 도 1b에는 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템의 전체적인 구성을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지 데이터 또는 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여, 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델, 객체 분석 모델 및 의도 분석 모델로부터 추출된 전시 대상 작품의 서로 다른 분석 특징(또는 서로 다른 분석 큐레이션 정보)을 동일한 텍스트 임베딩 공간 상으로 매핑되는 텍스트 임베딩 표현으로 추출하기 위한 서로 다른 분석 모델을 포함하고, 서로 다른 분석 모델로부터 추출된 서로 다른 특징에 관한 텍스트 임베딩 표현을 취합하기 위하여 concatenate layer로부터 개시되는 신경망 네트워크를 통하여 출력된 협의의 큐레이션과, 상기 협의의 큐레이션을 입력으로 하는 공간 추천 모델로부터 예측된 전시 공간에 관한 추천을 포함하여 광의의 큐레이션을 생성하기 위한 큐레이션 생성 시스템의 전체적인 구조를 설명하기 위한 도면이 도시되어 있다.

도 2에는 도 1에 도시된 색감 분석 모델로서, 전시 대상 작품을 형성하는 이미지로부터 색감 정보 내지는 컬러 톤(color tone) 정보를 표현하는 히스토그램(histogram) 내지는 히스토그램(histogram) 정보에 기반하는 색감 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품의 이미지를 형성하도록 채널 차원에서 합성되는 R,G,B의 3채널 영상에서 서로 다른 화소 값과 서로 다른 화소 값이 등장하는 화소 개수를 표시한 히스토그램을 예시적으로 보여주는 도면이 도시되어 있다.

도 3에는 도 1에 도시된 감성 분석 모델을 설명하기 위한 도면으로, 각각 전시 대상 작품을 형성하는 이미지 데이터를 입력으로 하는 이미지 기반의 감성 분석

모델과, 전시 대상 작품의 이미지 자체가 아닌, 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트 기반의 감성 분석 모델의 예측을 취합하기 위하여 concatenate layer로부터 개시되는 신경망 네트워크를 포함하는 감성 분석 모델의 전체적인 구성을 설명하기 위한 도면이 도시되어 있다.

도 4에는 도 1에 도시된 감성 분석 모델을 설명하기 위한 도면으로, 감성 분석 모델로부터 예측된 세분화된 감성 분류 클래스 또는 연속적인 감성 분류를 설명하기 위한 도면으로, Arousal 축과 Valence 축을 갖는 2차원의 평면 상에서, Arousal 축을 따라 양단의 active 및 passive와 Valence 축을 따라 양단의 negative와 positive로 하여 방사상으로 12 단계의 감성 분류를 포함하는 감성 분류의 공간을 예시적으로 보여주는 도면이 도시되어 있다.

도 5에는 도 3에 도시된 이미지 기반의 감성 분류 모델을 설명하기 위한 도면으로, CNN(convolution neural network) 계열의 네트워크를 포함하며, 예를 들어, convolution, pooling, linear fully connected layer를 포함하는 이미지 기반의 감성 분류 모델의 아키텍처를 설명하기 위한 도면이 도시되어 있다.

도 6 내지 도 8에는 도 3에 도시된 텍스트 기반의 감성 분류 모델을 설명하기 위한 도면으로, 도 6 및 도 7에는 상기 텍스트 기반의 감성 분류 모델에 적용될 수 있는 언어모델로서 GPT(Generative Pre-trained Transformer) 모델의 아키텍처를 개략적으로 보여주는 도면으로, 도 6은 다수의 디코더 블록을 쌓아 올린 GPT 모델의 아키텍처로서, 단어 시퀀스의 입력 순서에 따라 각각의 단어 표현을 앞서 입력된 단어와의 연관도에 따른 가중치의 합으로 산출하는 자연어 처리를 보여주는

도면이고, 도 7은 도 6에 도시된 디코더 블록의 보다 상세한 아키텍처를 보여주는 도면으로, 입력된 단어 시퀀스의 다음 단어를 예측하는 자연어 처리를 보여주는 도면이 도시되어 있으며, 도 8에는 현재 입력된 단어의 일 예시로서, 스페셜 토큰(S) 다음에 나올 단어를 예측하는 자연어 처리를 보여주는 도면으로, 스페셜 토큰(S)의 출력 값과 vocabulary를 형성하는 50,257 토큰 임베딩과의 전치 벡터 곱으로부터 소프트 맥스 함수를 취하여 예측 확률을 산출하는 자연어 처리를 보여주는 도면이 도시되어 있다.

도 9에는 도 3에 도시된 텍스트 기반의 감성 분류 모델을 설명하기 위한 도면으로, 도 6 내지 도 8에 도시된 바와 같은 텍스트 기반의 감성 분류 모델을 통하여 Text Prediction과 함께, Task classifier로서 입력 임베딩을 형성하는 스페셜 토큰으로 [Extract] 토큰에 해당되는 컨텍스트 표현을 추출하고, 예를 들어, Transformer 및/또는 Linear layer에 입력하여 감성 분석 내지는 감성 분류를 수행하는 텍스트 기반의 감성 분류 모델을 설명하기 위한 도면이 도시되어 있다.

도 10에는 도 1에 도시된 스타일 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지로부터 스타일 정보를 추출하고, 사전에 설정된 다수의 템플릿 이미지 각각으로부터 추출된 스타일 정보 사이의 유사도 분석을 통하여 가장 높은 유사도 스코어가 산출된 템플릿 이미지와 연계된 스타일에 관한 설명 또는 기술을 탐색하여 해당되는 전시 대상 작품의 스타일 분석의 특징 또는 스타일 분석의 큐레이션으로 생성하기 위한 스타일 분석 모델을 설명하기 위한 도면이 도시되어 있으며, 서로 다른 전시 대상 작품의 이미지와 템플릿 이미지 간의 유사도

분석을 위하여, 전시 대상 작품의 이미지와 템플릿 이미지의 서로 다른 이미지가 입력된 CNN 네트워크의 특정한 레이어에서 서로 다른 특징을 추출하기 위한 서로 다른 필터 또는 커널의 적용으로부터 산출된 서로 다른 특징(feature 1~4) 간의 상관관계(correlation)에 해당되는 스타일을 추출하는 것을 설명하기 위한 도면이 도시되어 있다.

도 11에는 도 1에 도시된 스타일 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품의 이미지와 템플릿 이미지의 서로 다른 이미지로부터 추출된 서로 다른 특징(feature 1~4) 간의 상관관계 내지는 스타일에 관한 Gram 매트릭스의 추출을 설명하기 위한 도면이 도시되어 있다.

도 12에는 도 1에 도시된 객체 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지 상에서 포착된 객체의 경계를 바운딩 박스로 예측하는 객체 인식(object detection)을 수행하기 위한 Faster R-CNN의 아키텍처를 포함하는 객체 분석 모델을 설명하기 위한 도면이 도시되어 있다.

도 13에는 도 1에 도시된 객체 분석 모델을 설명하기 위한 도면으로, 도 12에 도시된 ROI pooling을 설명하기 위한 도면이 도시되어 있다.

도 14에는 도 1에 도시된 객체 분석 모델을 설명하기 위한 도면으로, 도 12에 도시된 RPN(Region Proposal Network) 네트워크를 설명하기 위한 도면이 도시되어 있다.

도 15에는 도 1에 도시된 의도 분석 모델을 설명하기 위한 도면으로, bi-

directional language model인 BERT로부터 입력 임베딩을 형성하는 스페셜 토큰인 [CLS] 토큰의 표현 값에 해당되는 컨텍스트 표현에 기반하여 의도 분석 모델로 입력된 메타 데이터를 형성하는 단어 시퀀스, 단어 집합 또는 문장 집합에 대한 요약 생성을 구현하는 의도 분석 모델을 설명하기 위한 도면이 도시되어 있다.

도 16에는 도 1에 도시된 의도 분석 모델을 설명하기 위한 도면으로, forward language model인 GPT로부터 입력 임베딩을 형성하는 스페셜 토큰인 [Extract] 토큰의 표현 값에 해당되는 컨텍스트 표현에 기반하여 의도 분석 모델로 입력된 메타 데이터를 형성하는 단어 시퀀스, 단어 집합 또는 문장 집합에 대한 요약 생성을 구현하는 의도 분석 모델을 설명하기 위한 도면이 도시되어 있다.

도 17에는 도 1에 도시된 의도 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지 자체가 아닌, 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여 차원 축소시키기 위한 인코더(Encoder)로부터 입력된 메타 데이터의 전체적인 문맥(context) 정보를 포함하는 컨텍스트 표현을 추출하기 위한 시퀀스-투-시퀀스의 아키텍처를 포함하는 의도 분석 모델로서, 예를 들어, RNN 또는 LSTM 네트워크와 같은 시퀀스-투-시퀀스의 아키텍처를 포함하는 의도 분석 모델을 설명하기 위한 도면이 도시되어 있으며, 도 17에 도시된 도면은, 상기 인코더로부터 추출된 컨텍스트 표현을 입력으로 하여 차원 확장을 위한 디코더(Decoder)를 포함하는 인코더-디코더의 트랜스포머의 아키텍처로부터 번역의 태스크가 구현되는 것이 예시되어 있다.

도 18에는 도 1에 도시된 공간 추천 모델을 설명하기 위한 도면으로, 도 1에

도시된 서로 다른 분석 모델로부터 추출된 서로 다른 분석 특징에 기반한 협의의 큐레이션을 입력으로 하여, 입력된 협의의 큐레이션의 전체적인 문맥(context) 정보를 포함하는 컨텍스트 표현을 산출하도록 차원 축소를 위한 인코더(예를 들어, Embedding model)로부터 다차원의 텍스트 임베딩 공간 상으로 매핑되는 컨텍스트 표현에 관한 임베딩 표현을 산출하고 동일한 텍스트 임베딩 공간 상에서 각각의 클래스 공간에 관한 임베딩 표현과의 유클리드 거리에 기반하여 해당되는 전시 대상 작품에 관한 클래스 공간의 분류 내지는 전시 공간의 추천을 구현하는 공간 추천 모델을 설명하기 위한 도면이 도시되어 있다.

도 19에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, random noise z를 입력으로 하여 합성 데이터(생성 대상인 큐레이션)를 생성하기 위한 Generator G와 Generator G로부터 생성된 합성 데이터와 원본 데이터를 입력으로 하여 합성 데이터(Fake)와 원본 데이터(Real) 사이를 판별하기 위한 Discriminator D를 포함하는 GAN(Generative Adversarial Network)을 설명하기 위한 도면이며, 각각의 손실 함수(G loss, D loss)로부터 산출된 Gradient descent의 역전파(backpropagation) 알고리즘을 통한 Generator G와 Discriminator D의 학습을 설명하기 위한 도면이 도시되어 있다.

도 20에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, random noise z를 입력으로 하여 합성

데이터(생성 대상인 큐레이션)를 생성하기 위한 Generator G와 Generator G로부터 생성된 합성 데이터와 원본 데이터를 입력으로 하여 합성 데이터(Fake)와 원본 데이터(Real) 사이를 판별하기 위한 Discriminator D를 포함하는 GAN(Generative Adversarial Network)을 설명하기 위한 도면이 도시되어 있다.

도 21에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Time series based GAN에 적용될 수 있는 시퀀스(sequence) 모델의 예시로서, Time series based GAN에 포함될 수 있는 네트워크로서 RNN(Recurrent Neural Network)과 LSTM(Long Short Term Memory)의 구조를 예시적으로 보여주는 도면이 도시되어 있다.

도 22에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Time series based GAN에 적용될 수 있는 시퀀스(sequence) 모델의 예시로서, LSTM의 연쇄를 포함하는 Time series based GAN(Sequence GAN)의 Generator G의 예시를 보여주는 도면이 도시되어 있다.

도 23에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생

성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Generator G와 Discriminator D가 시퀀스 모델로서 LSTM(Long Short Term Memory)을 포함하거나, 또는 Generator G가 시퀀스 모델로서 LSTM(Long Short Term Memory)을 포함하고 Discriminator D가 CNN(Convolution Neural Network)을 포함하는 Sequence GAN의 아키텍처를 설명하기 위한 도면과, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Generator G가 MC search(Monte Carlo Search Tree)를 통한 보상(Reward)으로부터 정책 또는 정책 함수를 업-데이트 시키는 강화 학습 또는 강화 학습의 policy gradient를 설명하기 위한 도면이 도시되어 있다.

도 24에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Generator G가 latent vector z와 이전 시간 스텝의 출력으로부터 각각의 시간 스텝의 출력으로서 시계열의 합성 데이터(생성 대상인 큐레이션)를 생성하도록 LSTM(Long Short Term Memory)의 시퀀스 모델을 포함하는 구성을 설명하기 위한 도면이 도시되어 있다.

도 25에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Discriminator D가

Generator G로부터 출력되는 시계열의 합성 데이터(생성 대상인 큐레이션)를 입력으로 하여, 각각의 시간 스텝의 합성 데이터(생성 대상인 큐레이션)를 입력으로 하는 LSTM(Long Short Term Memory)의 시퀀스 모델의 결과에 대한 vote로부터 최종적인 판별을 추론하는 것을 설명하기 위한 도면이 도시되어 있다.

도 26에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN의 Discriminator D를 형성하는 CNN(Convolution Neural Network)을 설명하기 위한 도면이 도시되어 있다.

도 27에는 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템의 입력으로서 전시 대상 작품을 형성하는 이미지 자체가 아닌, 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 예시적으로 보여주는 도면이 도시되어 있다.

도 28에는 도 27에 예시된 메타 데이터로서 텍스트 데이터와 함께, 전시 대상 작품을 형성하는 이미지 데이터를 입력으로 하여, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템으로부터 생성된 큐레이션(전시 공간의 추천을 포함하는 광의의 큐레이션)을 예시적으로 보여주는 도면이 도시되어 있다.

【발명을 실시하기 위한 구체적인 내용】

【0065】 이하, 첨부된 도면을 참조하여, 본 발명의 바람직한 실시형태에 관한 큐레이션 생성 시스템에 대해 설명하기로 한다.

【0066】도 1a 및 도 1b에는 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템의 전체적인 구성을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지 데이터 또는 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여, 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델, 객체 분석 모델 및 의도 분석 모델로부터 추출된 전시 대상 작품의 서로 다른 분석 특징(또는 서로 다른 분석 큐레이션 정보)을 동일한 텍스트 임베딩 공간 상으로 매핑되는 텍스트 임베딩 표현으로 추출하기 위한 서로 다른 분석 모델을 포함하고, 서로 다른 분석 모델로부터 추출된 서로 다른 특징에 관한 텍스트 임베딩 표현을 취합하기 위하여 concatenate layer로부터 개시되는 신경망 네트워크를 통하여 출력된 협의의 큐레이션과, 상기 협의의 큐레이션을 입력으로 하는 공간 추천 모델로부터 예측된 전시 공간에 관한 추천을 포함하여 광의의 큐레이션을 생성하기 위한 큐레이션 생성 시스템의 전체적인 구조를 설명하기 위한 도면이 도시되어 있다.

【0067】도 2에는 도 1에 도시된 색감 분석 모델로서, 전시 대상 작품을 형성하는 이미지로부터 색감 정보 내지는 컬러 톤(color tone) 정보를 표현하는 히스토그램(histogram) 내지는 히스토그램(histogram) 정보에 기반하는 색감 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품의 이미지를 형성하도록 채널 차원에서 합성되는 R,G,B의 3채널 영상에서 서로 다른 화소 값과 서로 다른 화소 값이 등장하는 화소 개수를 표시한 히스토그램을 예시적으로 보여주는 도면이 도시되어 있다.

【0068】도 3에는 도 1에 도시된 감성 분석 모델을 설명하기 위한 도면으로, 각각 전시 대상 작품을 형성하는 이미지 데이터를 입력으로 하는 이미지 기반의 감성 분석 모델과, 전시 대상 작품의 이미지 자체가 아닌, 이미지에 부수되는 메타데이터로서 텍스트 데이터를 입력으로 하는 텍스트 기반의 감성 분석 모델의 예측을 취합하기 위하여 concatenate layer로부터 개시되는 신경망 네트워크를 포함하는 감성 분석 모델의 전체적인 구성을 설명하기 위한 도면이 도시되어 있다.

【0069】도 4에는 도 1에 도시된 감성 분석 모델을 설명하기 위한 도면으로, 감성 분석 모델로부터 예측된 세분화된 감성 분류 클래스 또는 연속적인 감성 분류를 설명하기 위한 도면으로, Arousal 축과 Valence 축을 갖는 2차원의 평면상에서, Arousal 축을 따라 양단의 active 및 passive와 Valence 축을 따라 양단의 negative와 positive로 하여 방사상으로 12 단계의 감성 분류를 포함하는 감성 분류의 공간을 예시적으로 보여주는 도면이 도시되어 있다.

【0070】도 5에는 도 3에 도시된 이미지 기반의 감성 분류 모델을 설명하기 위한 도면으로, CNN(convolution neural network) 계열의 네트워크를 포함하며, 예를 들어, convolution, pooling, linear fully connected layer를 포함하는 이미지 기반의 감성 분류 모델의 아키텍처를 설명하기 위한 도면이 도시되어 있다.

【0071】도 6 내지 도 8에는 도 3에 도시된 텍스트 기반의 감성 분류 모델을 설명하기 위한 도면으로, 도 6 및 도 7에는 상기 텍스트 기반의 감성 분류 모델에 적용될 수 있는 언어모델로서 GPT(Generative Pre-trained Transformer) 모델의 아키텍처를 개략적으로 보여주는 도면으로, 도 6은 다수의 디코더 블록을 쌓아 올린

GPT 모델의 아키텍처로서, 단어 시퀀스의 입력 순서에 따라 각각의 단어 표현을 앞서 입력된 단어와의 연관도에 따른 가중치의 합으로 산출하는 자연어 처리를 보여주는 도면이고, 도 7은 도 6에 도시된 디코더 블록의 보다 상세한 아키텍처를 보여주는 도면으로, 입력된 단어 시퀀스의 다음 단어를 예측하는 자연어 처리를 보여주는 도면이 도시되어 있으며, 도 8에는 현재 입력된 단어의 일 예시로서, 스페셜 토큰(S) 다음에 나올 단어를 예측하는 자연어 처리를 보여주는 도면으로, 스페셜 토큰(S)의 출력 값과 vocabulary를 형성하는 50,257 토큰 임베딩과의 전치 벡터 곱으로부터 소프트 맥스 함수를 취하여 예측 확률을 산출하는 자연어 처리를 보여주는 도면이 도시되어 있다.

【0072】 도 9에는 도 3에 도시된 텍스트 기반의 감성 분류 모델을 설명하기 위한 도면으로, 도 6 내지 도 8에 도시된 바와 같은 텍스트 기반의 감성 분류 모델을 통하여 Text Prediction과 함께, Task classifier로서 입력 임베딩을 형성하는 스페셜 토큰으로 [Extract] 토큰에 해당되는 컨텍스트 표현을 추출하고, 예를 들어, Transformer 및/또는 Linear layer에 입력하여 감성 분석 내지는 감성 분류를 수행하는 텍스트 기반의 감성 분류 모델을 설명하기 위한 도면이 도시되어 있다.

【0073】 도 10에는 도 1에 도시된 스타일 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지로부터 스타일 정보를 추출하고, 사전에 설정된 다수의 템플릿 이미지 각각으로부터 추출된 스타일 정보 사이의 유사도 분석을 통하여 가장 높은 유사도 스코어가 산출된 템플릿 이미지와 연계된 스타일에 관

한 설명 또는 기술을 탐색하여 해당되는 전시 대상 작품의 스타일 분석의 특징 또는 스타일 분석의 큐레이션으로 생성하기 위한 스타일 분석 모델을 설명하기 위한 도면이 도시되어 있으며, 서로 다른 전시 대상 작품의 이미지와 템플릿 이미지 간의 유사도 분석을 위하여, 전시 대상 작품의 이미지와 템플릿 이미지의 서로 다른 이미지가 입력된 CNN 네트워크의 특정한 레이어에서 서로 다른 특징을 추출하기 위한 서로 다른 필터 또는 커널의 적용으로부터 산출된 서로 다른 특징(feature 1~4) 간의 상관관계(correlation)에 해당되는 스타일을 추출하는 것을 설명하기 위한 도면이 도시되어 있다.

【0074】도 11에는 도 1에 도시된 스타일 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품의 이미지와 템플릿 이미지의 서로 다른 이미지로부터 추출된 서로 다른 특징(feature 1~4) 간의 상관관계 내지는 스타일에 관한 Gram 매트릭스의 추출을 설명하기 위한 도면이 도시되어 있다.

【0075】도 12에는 도 1에 도시된 객체 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지 상에서 포착된 객체의 경계를 바운딩 박스로 예측하는 객체 인식(object detection)을 수행하기 위한 Faster R-CNN의 아키텍처를 포함하는 객체 분석 모델을 설명하기 위한 도면이 도시되어 있다.

【0076】도 13에는 도 1에 도시된 객체 분석 모델을 설명하기 위한 도면으로, 도 12에 도시된 ROI pooling을 설명하기 위한 도면이 도시되어 있다.

【0077】도 14에는 도 1에 도시된 객체 분석 모델을 설명하기 위한 도면으로, 도 12에 도시된 RPN(Region Proposal Network) 네트워크를 설명하기 위한 도면이 도시되어 있다.

【0078】도 15에는 도 1에 도시된 의도 분석 모델을 설명하기 위한 도면으로, bi-directional language model인 BERT로부터 입력 임베딩을 형성하는 스페셜 토큰인 [CLS] 토큰의 표현 값에 해당되는 컨텍스트 표현에 기반하여 의도 분석 모델로 입력된 메타 데이터를 형성하는 단어 시퀀스, 단어 집합 또는 문장 집합에 대한 요약 생성을 구현하는 의도 분석 모델을 설명하기 위한 도면이 도시되어 있다.

【0079】도 16에는 도 1에 도시된 의도 분석 모델을 설명하기 위한 도면으로, forward language model인 GPT로부터 입력 임베딩을 형성하는 스페셜 토큰인 [Extract] 토큰의 표현 값에 해당되는 컨텍스트 표현에 기반하여 의도 분석 모델로 입력된 메타 데이터를 형성하는 단어 시퀀스, 단어 집합 또는 문장 집합에 대한 요약 생성을 구현하는 의도 분석 모델을 설명하기 위한 도면이 도시되어 있다.

【0080】도 17에는 도 1에 도시된 의도 분석 모델을 설명하기 위한 도면으로, 전시 대상 작품을 형성하는 이미지 자체가 아닌, 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여 차원 축소시키기 위한 인코더(Encoder)로부터 입력된 메타 데이터의 전체적인 문맥(context) 정보를 포함하는 컨텍스트 표현을 추출하기 위한 시퀀스-투-시퀀스의 아키텍처를 포함하는 의도 분

석 모델로서, 예를 들어, RNN 또는 LSTM 네트워크와 같은 시퀀스-투-시퀀스의 아키텍처를 포함하는 의도 분석 모델을 설명하기 위한 도면이 도시되어 있으며, 도 17에 도시된 도면은, 상기 인코더로부터 추출된 컨텍스트 표현을 입력으로 하여 차원 확장을 위한 디코더(Decoder)를 포함하는 인코더-디코더의 트랜스포머의 아키텍처로부터 번역의 태스크가 구현되는 것이 예시되어 있다.

【0081】도 18에는 도 1에 도시된 공간 추천 모델을 설명하기 위한 도면으로, 도 1에 도시된 서로 다른 분석 모델로부터 추출된 서로 다른 분석 특징에 기반한 협의의 큐레이션을 입력으로 하여, 입력된 협의의 큐레이션의 전체적인 문맥(context) 정보를 포함하는 컨텍스트 표현을 산출하도록 차원 축소를 위한 인코더(예를 들어, Embedding model)로부터 다차원의 텍스트 임베딩 공간 상으로 매핑되는 컨텍스트 표현에 관한 임베딩 표현을 산출하고 동일한 텍스트 임베딩 공간 상에서 각각의 클래스 공간에 관한 임베딩 표현과의 유클리드 거리에 기반하여 해당되는 전시 대상 작품에 관한 클래스 공간의 분류 내지는 전시 공간의 추천을 구현하는 공간 추천 모델을 설명하기 위한 도면이 도시되어 있다.

【0082】본 발명의 일 실시형태에서 큐레이션(curation)은 미술관, 박물관 등에 전시되는 작품을 기획하고 설명해주는 큐레이터(curator)에서 파생된 개념으로 이해될 수 있으며, 예를 들어, 큐레이션은 큐레이터와 유사하게, 전시 공간에 대한 추천과 전시 대상 작품에 대한 설명이나 서술을 제공하여 전시 대상 작품에 대한 이해를 돕고 가치를 부여하도록 또는 이를 통하여 전시 대상 작품을 소비할 수 있도록 돕기 위한 목적으로, 텍스트 시퀀스 형태로 제공되는 전시 대상 작품에

대한 설명 내지는 서술을 의미할 수 있다. 즉, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 1) 전시 대상 작품에 관한 컨텐츠 데이터로부터 전시 대상 작품에 대한 설명이나 서술(narration)을 제공할 수 있으며, 또한 2) 전시 대상 작품으로부터 전시 공간을 추천해줄 수 있다. 이와 같이, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은 1) 전시 대상 작품에 대한 설명이나 서술로서 2) 전시 대상 작품과 매칭되는 전시 공간을 추천해줄 수 있으며, 이와 같은 실시형태에서, 상기 2) 전시 대상 작품과 매칭되는 전시 공간에 대한 추천은 1) 전시 대상 작품에 대한 설명이나 서술로서 하나의 텍스트 시퀀스의 형태로 제공될 수 있다. 본 발명의 일 실시형태에서 텍스트 시퀀스란 하나의 문장을 제한적으로 의미하기 보다는, 예를 들어, 다수의 문장을 포함하는 문장 집합을 포괄적으로 의미할 수 있으며, 이런 의미에서 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템의 출력으로서 텍스트 시퀀스란 하나 이상 다수의 문장을 포함할 수 있다.

【0083】 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 서로 다른 멀티-모달(multi-modal)의 컨텐츠 데이터를 입력으로 하여, 텍스트 시퀀스 형태의 설명 내지는 서술을 생성할 수 있으며, 예를 들어, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 전시 대상 작품에 대한 컨텐츠 데이터로서 텍스트 데이터와 이미지 데이터를 포함하는 멀티-모달의 컨텐츠 데이터를 입력으로 하여 해당되는 전시 대상 작품에 대한 설명 내지는 서술을 생성할 수 있으며, 이때, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은 전시 대상 작품으로부터 다양한 feature를 추출해내고, 추출된 서로 다른 feature를 텍스트 시퀀스 형태로 취합하

여 제공할 수 있다.

【0084】본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 전시 대상 작품의 컨텐츠 데이터를 입력으로 하여, 색감 분석, 감성 분석, 스타일 분석, 객체 분석, 의도 분석 등을 구현하면서 전시 대상 작품을 형성하는 이미지 상에 등장하는 화소의 분포 또는 화소의 개수에 기반하는 전체 이미지 상의 색감이나 컬러 톤에 관한 feature와, Arousal 축과 Valence 축을 갖는 2차원의 평면 상에서 Arousal 축을 따라 양단의 active 및 passive와, Valence 축을 따라 양단의 negative와 positive로 하여 방사상으로 12 단계의 감성 분류를 포함하는 감성 분류 공간 상에서 연속적인 감성 분류 공간 상에서의 매핑에 기반한 감성에 관한 feature와, 사전에 수집된 템플릿 이미지와의 유사도 분석으로부터 가장 유사한 스타일의 템플릿 이미지로부터 추출된 스타일 정보와, 전시 대상 작품을 형성하는 이미지 상으로부터 객체 인식(object detection)을 통하여 인식된 객체에 관한 feature, 전시 대상 작품에 대해 작품을 기획한 작가의 주관적인 의견으로서, 작가의 작업 의도, 작가의 창작 의도 또는 작가의 철학 등이 포함된 작가 노트나, 작가 외의 전문가로부터의 작품 평, 작가 등이 명명한 작품 명으로부터 추출된 작품 의도에 관한 feature 등을 취합하여, 텍스트 시퀀스 형태로 전시 대상 작품에 관한 설명 내지는 서술이 생성될 수 있다. 본 발명의 일 실시형태에서는 앞서 설명된 바와 같은 서로 다른 유형의 분석을 수행하기 위한 서로 다른 네트워크 또는 알고리즘을 포함할 수 있으며, 보다 구체적으로, 색감 분석을 위한 색감 분석 모델, 감성 분석을 위한 감성 분석 모델, 스타일 분석을 위한 스타일 분석 모델, 객체 분석을 위한 객체 분석 모

델, 의도 분석을 위한 의도 분석 모델을 포함하고 이들 서로 다른 분석 모델로부터 추출된 feature들을 취합하여 분석 대상 내지는 전시 대상 작품에 대한 설명 내지는 서술을 생성하기 위한 concatenate 네트워크를 포함할 수 있다. 본 발명의 일 실시형태에서, 각각의 분석 모델은 신경망 네트워크로 구현되거나 또는 사전에 설정된 로직에 따라 프로세스를 수행하기 위한 알고리즘으로 구현될 수도 있다.

【0085】본 발명의 일 실시형태에서, 상기 색감 분석 모델은, 분석 대상 내지는 전시 대상 작품을 형성하는 이미지로부터 색감 정보 내지는 컬러 톤 정보를 추출할 수 있으며, 예를 들어, 분석 대상 내지는 전시 대상 작품을 형성하는 이미지로부터 컬러 톤(color tone) 내지는 색감 정보를 표현하는 히스토그램(histogram) 내지는 히스토그램 정보에 기반하여 전체 이미지 상에서 표현되는 색감 내지는 컬러 톤(color tone) 정보를 추출할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 색감 분석 모델은 분석 대상 내지는 전시 대상 작품을 형성하는 이미지를 형성하도록 합성되는 3채널 이미지 각각에 대해 화소 값 별로 등장하는 화소 개수를 표현하는 히스토그램(histogram) 내지는 히스토그램 정보에 기반하여 이미지 상에서 표현되는 색감 내지는 컬러 톤 정보를 추출할 수 있으며, 본 발명의 일 실시형태에서는 3채널 이미지 각각으로부터 추출된 제1 내지 제3 히스토그램 내지는 제1 내지 제3 히스토그램 정보에 기반하여 이미지 상에서 표현되는 색감 내지는 컬러 톤 정보를 추출할 수 있다. 예를 들어, 분석 대상 내지는 전시 대상 작품을 형성하는 이미지의 일 특징(feature)으로서 컬러 톤(color tone) 내지는 색감이란 이미지를 표현하는 색상의 배분 정보로 규정될 수 있으며, 색상의 배분 정

보로서 서로 다른 색상을 표현하도록 서로에 대해 합성되는 R, G, B 3채널 이미지(또는 Y, Cb, Cr 3채널 이미지) 상에서 화소 값 별로 해당되는 화소 값이 등장하는 화수 개수를 표현하는 히스토그램 내지는 히스토그램 정보를 추출할 수 있다. 본 발명의 일 실시형태에서는 분석 대상 내지는 전시 대상 작품을 형성하는 이미지로부터 컬러 톤 내지는 색감 정보를 함축한 히스토그램 내지는 히스토그램 정보를 추출할 수 있으며, 히스토그램 내지는 히스토그램 정보에 기반하여 이미지의 전반적인 컬러 톤 내지는 색감 정보의 특징을 추출할 수 있다.

【0086】본 발명의 일 실시형태에서, 전시 대상 작품을 형성하는 이미지로부터 컬러 톤 내지는 색감 정보를 표현하는 히스토그램은, 이미지 상에서 서로 다른 색상을 표현하도록 서로에 대해 합성되는 서로 다른 3채널의 이미지로서, R,G,B 컬러 스페이스 상에서 R,G,B 3채널 이미지 또는 Y,Cb,Cr 컬러 스페이스 상에서 Y,Cb,Cr 3채널 이미지 상에서 화소 값 별로 해당되는 화소 값이 3채널 이미지 각각에서 등장하는 화소 개수를 표현할 수 있다. 예를 들어, 전시 대상 작품을 형성하는 이미지의 일 특징(feature)으로 컬러 톤 내지는 색감 정보는, 전시 대상 작품의 이미지를 형성하는 3채널 이미지 각각으로부터 추출된 제1 내지 제3 히스토그램 내지는 제1 내지 제3 히스토그램 정보에 기반하여, 추출될 수 있다.

【0087】본 발명의 일 실시형태에서, 상기 감성 분석 모델은, 분석 대상 내지는 전시 대상 작품의 컨텐츠 데이터를 입력으로 하여, 전시 대상 작품에 표현된 감성을 서로 다른 감성 클래스로 분류할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, 상기 감성 분석 모델은 전시 대상 작품의 컨텐츠 데이터로서 텍스트 데

이터와 이미지 데이터의 멀티-모달의 데이터로부터 각각 감성 분류를 구현할 수 있으며, 전시 대상 작품의 컨텐츠 데이터로서 텍스트 데이터를 입력으로 하는 텍스트 기반의 감성 분석 모델과 전시 대상 작품의 이미지 데이터로서 이미지 데이터를 입력으로 하는 이미지 기반의 감성 분석 모델을 포함할 수 있으며, 상기 텍스트 기반의 감성 분석 모델과 이미지 기반의 감성 분석 모델의 예측을 취합하여 전시 대상 작품을 형성하는 이미지 상에 표현된 감성을 예측할 수 있다.

【0088】 본 발명의 일 실시형태에서는, 전시 대상 작품의 컨텐츠 데이터로서 이미지 데이터와 텍스트 데이터의 멀티 모달(multi-modal)의 데이터로부터 감성 분석(emotion recognition)를 수행할 수 있으며, 예를 들어, 상기 감성 분석은 사전에 설정된 감성 클래스에 속할 확률을 예측하기 위한 감성 분류를 구현할 수 있으며, 이때, 상기 감성 분석 내지는 감성 분류는 이산적으로 분류된 서로 다른 감성 클래스로서, anger, disgust, fear, happiness, sadness, surprise와 같은 6 단계의 감성 클래스(또는 AU, action unit)로 분류될 수 있으며, 사전에 설정된 이산적인 감성 클래스에 속할 확률을 예측하기 보다는, 예를 들어, 연속적인 감성 분류로서, Arousal 축과 valence 축을 갖는 2차원 평면 상에서 Arousal 축을 따라 양단의 active 및 passive와, Valence 축을 따라 양단의 negative와 positive로 하여, 방사상으로 배열된 12 단계의 감성 분류를 포함하는 감성 분류 공간 상에서의 매핑에 기반한 감성 분석을 구현할 수도 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 감성 분석 모델은 분석 대상 작품의 컨텐츠 데이터로서 이미지 데이터 및 텍스트 데이터를 포함하는 멀티-모달의 데이터를 입력으로 하여, 분석 대상 작품의 이

미지 상에 표현된 감성을 분석할 수 있으며, 이미지 데이터를 입력으로 하는 이미지 기반의 감성 분석 모델과 텍스트 기반의 감성 분석 모델의 예측을 취합하여 이미지 상에 표현된 감성을, 상기 감성 분류의 공간 상에 매핑할 수 있으며, 감성 분류의 공간 상으로 매핑된 전시 대상 작품의 감성에 관한 임베딩 표현과의 유클리드 거리에 근거하여 해당되는 전시 대상 작품의 감성을 분석할 수 있으며, 예를 들어, 전시 대상 작품의 감성에 관하여 상기 감성 분류의 공간 상에 매핑된 임베딩 표현과의 유클리드 거리에 따라 일 군의 임베딩 표현을 카테고라이징 하는 감성에 관한 표현(representation)으로부터 해당되는 전시 대상 작품 상에 표현된 감성을 예측하거나 또는 상기 감성 분류의 공간 상에 매핑된 다수의 임베딩 표현을 일 군의 군집으로 카테고라이징 하는 군집의 중심 사이의 거리에 따라 이산적인 형태의 감성 분류의 클래스에 의하지 않고 보다 세분화된 감성 분류 또는 연속적인 감성 분류의 공간으로 매핑되도록 각각의 클래스(예를 들어, 12 단계의 감성 분류)에 속할 것으로 예측된 확률에 따라 상기 감성 분류의 공간으로 매핑된 전시 대상 작품의 감성에 관한 임베딩 표현으로부터 세분화된 또는 연속적인 감성 분류 내지는 감성 분석을 구현할 수 있다. 본 발명의 일 실시형태에서 전시 대상 작품의 일 특징(feature)으로서 감성 분석은 전시 대상 작품에 관한 설명 또는 서술로서, 예를 들어, 전시 대상 작품의 예술적인 이해를 돕고 예술적인 가치를 부여하기 위한 것으로, 감성 분석 모델로부터 추출된 전시 대상 작품에 관하여 예측된 감성은, 전시 대상 작품의 예술적인 이해를 돕고 예술적인 가치를 부여하기 위한 설명 또는 서술의 주요한 일부를 형성하는 것으로 이해될 수 있으며, 이러한 맥락에서 본 발명의

일 실시형태에서 감성 분석 모델은 보다 세분화된 감성 분류의 클래스 또는 연속적인 감성 분류를 구현하면서 보다 세밀하고 정교한 감성 분석 내지는 감성 분류를 구현할 수 있다.

【0089】 본 발명의 일 실시형태에 따른 감성 분석 모델은, 전시 대상 작품을 형성하는 이미지 자체를 입력으로 하는 이미지 기반의 감성 분석 모델과, 전시 대상 작품을 형성하는 이미지 자체가 아닌, 전시 대상 작품에 부수되는 메타 데이터로서, 전시 대상 작품을 기획한 작가로부터 제공된 주관적인 의견으로서 작가의 작업 의도, 작가의 창작 의도, 작가의 철학 등이 포함된 작가 노트, 또는 전시 대상 작품에 대한 전문가의 작품 평, 전시 대상 작품을 기획한 작자(또는 작가 외의 전문가)가 명명한 작품 명을 포함하는 작품 캡션 정보(예를 들어, 작품 명, 크기, 매체 등)에 관한 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트 기반의 감성 분석 모델을 포함할 수 있다.

【0090】 예를 들어, 본 발명의 일 실시형태에서, 이미지 기반의 감성 분석 모델은 CNN(convolution neural network) 계열의 네트워크를 포함할 수 있으며, 상기 CNN 계열의 네트워크는 다수의 학습 데이터로서 전시 대상 작품을 형성하는 이미지와 각각의 이미지에 대한 타겟 레이블로서 사전에 지정된 감성 분류의 클래스 중에서 어느 하나의 클래스로 주어진 타겟 레이블로부터 학습된 가중치를 포함할 수 있다.

【0091】 예를 들어, 상기 이미지 기반의 감성 분석 모델에서는, 전시 대상 작품의 이미지 자체를 입력으로 하는 CNN 계열의 네트워크를 포함할 수 있으며, 이

미지 기반의 감성 분석 모델로서, CNN 계열의 네트워크는, convolution, pooling 및 dropout(미도시)을 포함하는 각각의 레이어가 적층된 아키텍처를 포함할 수 있으며, linear fully connected layer를 통과하여, 입력된 이미지가 각각의 감성 클래스에 속할 확률을 예측할 수 있다.

【0092】상기 텍스트 기반의 감성 분석 모델은 전시 대상 작품의 이미지 자체가 아닌, 예를 들어, 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 언어 모델의 네트워크를 포함할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, 상기 언어 모델로서 forward language model인 GPT 모델에서는 어텐션 블록(attention block)을 다수로 쌓아 올린 멀티 헤드 어텐션(multi-head attention)과 피드 포워드 네트워크(feedforwad network)를 포함하는 디코더 블록을 누적하여 다수로 쌓아 올리고, 각 디코더 블록의 결과 값이 다음으로 연결된 디코더 블록의 입력으로 들어가도록 다수의 디코더 블록이 연결될 수 있으며, 최후의 디코더 블록의 결과 값으로부터 최초의 디코더 블록으로 입력된 메타 데이터로서 메타 데이터를 형성하는 텍스트 또는 메타 데이터를 형성하는 텍스트에 관한 임베딩 표현에 대한 전체적인 문맥 정보를 포함하는 컨텍스트 표현(context vector)가 산출될 수 있다. 상기 디코더 블록은 멀티 헤드 어텐션(multi-head attention)과 피드 포워드 네트워크(feedforwad network)를 포함할 수 있으며, 이외에 skip connection을 위한 Add와 임베딩 표현(또는 토큰)에 대한 정규화(예를 들어, 임베딩 표현 또는 토큰을 0~1 사이로 정규화)를 위한 Layer Normalization(Layer Norm)을 포함할 수 있다.

【0093】본 발명의 일 실시형태에서 텍스트 기반의 감성 분석 모델은 전시 대상 작품에 관하여 입력된 메타 데이터로서 텍스트의 메타 데이터를 입력으로 자연어 처리를 구현할 수 있으며, forward language model인 GPT 모델에서는 입력된 텍스트 이후에 등장할 다음 단어를 예측할 수 있으며(Text Prediction), 본 발명의 일 실시형태에서 전시 대상 작품에 관한 메타 데이터로부터 전시 대상 작품에 표현된 감성을 예측하기 위한 감성 분석 내지는 감성 분류를 구현할 수 있다(Task Classification). 예를 들어, 상기 텍스트 기반의 감성 분석 모델은 입력된 메타 데이터의 전체적인 문맥(context) 정보를 포함하는 컨텍스트 표현(context vector)을 산출할 수 있으며, 산출된 컨텍스트 표현을 Transformer 및/또는 Linear layer에 입력하여 감성 분석 내지는 감성 분류를 수행할 수 있다. 예를 들어, 본 발명의 일 실시형태에 따른 텍스트 기반의 감성 분석 모델로서 언어 모델에는 토큰 임베딩(token embedding), 위치 임베딩(positional embedding), 세그먼트 임베딩(segment embedding)이 합산된 입력 임베딩이 입력될 수 있으며, 상기 입력 임베딩은 문장 시퀀스의 시작에 해당되는 스페셜 토큰(예를 들어, Start token)으로부터 문장 시퀀스의 종료 이후의 스페셜 토큰(예를 들어, Extract token)을 포함할 수 있으며, 최후의 디코더 블록의 Extract 토큰에 해당되는 컨텍스트 표현을 추출하여 Transformer 및/또는 Linear layer에 입력하여 감성 분석 내지는 감성 분류를 수행할 수 있으며, 상기 Transformer는 상기 감성 분석 모델을 형성하는 언어 모델로부터 추출된 컨텍스트 표현을 입력으로 하여 차원 축소시키기 위한 인코더(수축 경로, contracting path)와 인코더를 통하여 차원 축소된(또는 인코딩된) 컨텍스트

표현을 입력으로 하여 차원 확장시키기 위한 디코더(확장 경로, expanding path)를 포함하는 인코더-디코더 아키텍처를 포함하거나 또는 인코더 및 디코더 중에서 어느 하나를 포함할 수 있으며, 상기 Linear layer에는, 상기 언어 모델로부터 추출된 컨텍스트 표현이나 또는 상기 Transformer의 인코더로부터 추출된 인코딩된 컨텍스트 표현이나 또는 상기 Transformer의 인코더 및 디코더를 통하여 추출된 디코딩된 컨텍스트 표현이 입력될 수 있으며, 상기 Linear layer로부터 사전에 설정된 각각의 감성 클래스에 속할 확률이 산출될 수 있으며, 예를 들어, 상기 Linear layer로부터 산출되는 각각의 감성 클래스에 속할 확률로부터 해당되는 전시 대상 작품의 감성에 관한 임베딩은 12단계로 분류된 감성 분류의 공간 상으로 매핑될 수 있으며, 감성 분류의 공간 상으로 매핑된 전시 대상 작품의 임베딩으로부터 유클리드 거리에 기반하여 감성 분석 내지는 감성 분류가 구현될 수 있다.

【0094】본 발명의 일 실시형태에서, 전시 대상 작품의 이미지 자체를 입력으로 하여 전시 대상 작품 상에 표현된 감성에 관한 예측 내지는 임베딩 표현을 출력하기 위한 이미지 기반의 감성 분석의 데이터 흐름과 전시 대상 작품의 이미지 자체가 아닌, 이미지에 수반된 메타 데이터를 입력으로 하여 전시 대상 작품 상에 표현된 감성에 관한 예측 내지는 임베딩 표현을 출력하기 위한 텍스트 기반의 감성 분석의 데이터 흐름은, concatenate layer를 통하여 합류되면서 이후의 layer들을 통하여 이미지 기반의 감성 분석과 텍스트 기반의 감성 분석의 예측이 서로 취합될 수 있으며, 보다 구체적으로, 이미지 기반의 감성 분석의 출력과 텍스트 기반의 감성 분석의 출력을 concatenate 및 flatten 시키기 위한 layer, linear fully

connected layer, 임베딩 표현 내지는 토큰의 정규화를 위한 normalization layer 등을 통하여 이미지 기반의 감성 분석과 텍스트 기반의 감성 분석이 서로 종합적으로 취합된 결과로서 입력된 전시 대상 작품의 감성에 관한 특징(feature)을 추출할 수 있다.

【0095】 상기 스타일 분석 모델은 전시 대상 작품으로부터 스타일 정보를 추출할 수 있으며, 본 명세서를 통하여 스타일(style) 이란 전시 대상 작품을 형성하는 이미지 상에서 표현되는 서로 다른 특징(feature) 간의 상관관계(correlation)를 의미할 수 있고, 예를 들어, CNN 계열의 네트워크에서 이미지 상에 표현된 서로 다른 특징(feature)을 추출하기 위한 서로 다른 커널(kernel 또는 필터)의 적용으로부터 산출되는 서로 다른 특징(feature) 간의 상관관계에 해당되는 스타일 정보를 추출할 수 있다. 예를 들어, 본 발명의 일 실시형태에서 상기 전시 대상 작품을 형성하는 이미지로부터 추출된 서로 다른 특징(feature) 간의 상관관계를 표현한 Gram 매트릭스(Gram matrix)를 산출할 수 있다.

【0096】 본 발명의 일 실시형태에서, 스타일(style)이란 이미지를 형성하는 서로 다른 특징(feature) 간의 상관관계(correlation)로 규정될 수 있으며, 예를 들어, 이미지를 형성하는 서로 다른 특징(feature)이란, 선, 면과 같은 윤곽(profile)을 포함하는 geometry나 컬러 톤(color tone)과 같은 이미지를 형성하는 서로 다른 특징(feature)을 포함할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 스타일(style)이란 전시 대상 작품을 형성하는 이미지로부터 추출된 서로 다른 특징 1(feature 1) 및 특징 2(feature 2)에서, 특징 1에서 높은 activation 값을

가질 때, 특징 2에서도 높은 activation 값을 갖는다는 것을 의미할 수 있으며, 예를 들어, 특징 1 및 특징 2의 activation 값을 표현하는 벡터를 내적(dot product)하여 스타일 정보를 함축한 Gram 매트릭스를 산출할 수 있으며, 예를 들어, 전시 대상 작품을 형성하는 이미지와 사전에 설이미지로부터 스타일 정보를 함축한 Gram 매트릭스를 산출할 수 있고, 이와 같이, 전시 대상 작품을 형성하는 이미지와 템플릿 이미지로부터 추출된 Gram 매트릭스 사이의 유사도 분석에 기반하여, 전시 대상 작품의 스타일 정보와 템플릿 이미지의 스타일 정보 사이의 유사도를 분석함으로써, 다시 말하면, 본 발명의 일 실시형태에서 전시 대상 작품의 스타일 정보를 함축한 Gram 매트릭스와, 템플릿 이미지의 스타일 정보를 함축한 Gram 매트릭스 사이의 유사도를 분석함으로써, 즉, 전시 대상 작품을 형성하는 이미지로부터 추출된 스타일 정보를 함축한 Gram 매트릭스와 사전에 설정된 템플릿 이미지로부터 추출된 스타일 정보를 함축한 Gram 매트릭스 사이의 유사도 분석으로부터 전시 대상 작품의 이미지와 템플릿 이미지 사이의 스타일 유사도에 관한 스코어를 산출할 수 있다.

【0097】본 발명의 일 실시형태에서, 상기 스타일 분석 모델은 전시 대상 작품으로부터 스타일 정보를 함축한 Gram 매트릭스 자체로부터 해당되는 전시 대상 작품을 형성하는 스타일에 관한 feature가 추출된다고 할 수 있으나, 예를 들어, 앞서 설명된 바와 같이, 전시 대상 작품을 형성하는 이미지로부터 서로 다른 특징(feature)들을 추출하기 위한 색감 분석 모델, 감성 분석 모델, 객체 분석 모델, 의도 분석 모델과 달리, 전시 대상 작품을 형성하는 이미지 자체로부터 추출된 특

징(feature)에 기반하여 해당되는 전시 대상 작품에 대한 설명 또는 서술이 제공된다기 보다는, 사전에 설정된 템플릿 이미지와의 유사도 분석에 기반하여 해당되는 전시 대상 작품에 대한 설명 또는 서술이 생성된다고 할 수 있으며, 예를 들어, 전시 대상 작품으로부터 스타일 정보를 함축한 Gram 매트릭스로부터 직접적으로 해당되는 전시 대상 작품에 관한 스타일 정보를 포함하는 설명 또는 서술이 생성된다기보다는, 예를 들어, 사전에 설정된 템플릿 이미지로부터 스타일 정보를 함축한 Gram 매트릭스를 추출하고, 사전에 설정된 다수의 템플릿 이미지 각각으로부터 스타일 정보를 추출한 Gram 매트릭스와, 전시 대상 작품을 형성하는 이미지로부터 스타일 정보를 추출한 Gram 매트릭스 사이에서 가장 높은 유사도 스코어로 매칭된 템플릿 이미지와 연계되어 저장된 설명 내지는 서술을 탐색하여 탐색된 설명 내지는 서술로부터 해당되는 전시 대상 작품에 대한 스타일에 관한 설명 내지는 서술(예를 들어, 스타일 분석 큐레이션)이 생성될 수 있으며, 예를 들어, 해당되는 전시 대상 작품과 가장 높은 유사도 스코어로 매칭된 템플릿 이미지와 연계되어 저장된 설명 내지는 서술 자체가 해당되는 전시 대상 작품의 스타일에 관한 설명 내지는 서술로서 스타일 분석 큐레이션으로 생성되거나 또는 템플릿 이미지와 연계되어 저장된 설명 내지는 서술에 대한 편집으로부터 전시 대상 작품의 스타일에 관한 설명 내지는 서술로서 스타일 분석 큐레이션이 생성될 수도 있다.

【0098】본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 전시 대상 작품을 형성하는 이미지의 서로 다른 특징(feature)의 추출을 위한 서로 다른 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델, 객체 분석 모델, 의도 분석 모델로

부터 해당되는 전시 대상 작품의 서로 다른 특징(feature)을 추출하고 추출된 서로 다른 특징(feature) 각각에 기반하여 생성된 색감 분석 큐레이션, 감성 분석 큐레이션, 스타일 분석 큐레이션, 객체 분석 큐레이션, 의도 분석 큐레이션이 종합적으로 취합된 형태로 해당되는 전시 대상 작품에 관한 전체적인 큐레이션이 생성될 수 있다. 다시 말하면, 본 발명의 일 실시형태에서, 상기 스타일 분석 모델에서도, 스타일에 관한 특징(feature)과 다른 특징(feature)을 추출하기 위한 다른 모델에서와 유사하게, 해당되는 전시 대상 작품을 형성하는 이미지로부터 추출된 스타일 정보, 보다 구체적으로, 스타일 정보를 함축한 Gram 매트릭스를 추출하고, 추출된 스타일 정보 내지는 스타일 정보를 함축한 Gram 매트릭스 정보로부터 해당되는 전시 대상 작품에 관한 스타일 분석 큐레이션을 생성하는 것으로 이해될 수 있으나, 다른 모델에서와 달리, 사전에 설정된 템플릿 이미지와의 유사도 분석을 통하여 해당되는 전시 대상 작품과 가장 높은 유사도 스코어가 산출된 템플릿 이미지를 선정하고, 이와 같이 선정된 또는 매칭된 템플릿 이미지와 연계되어 저장된 설명 내지는 서술 자체가 해당되는 전시 대상 작품에 관한 스타일 분석 큐레이션으로 생성되거나 또는 매칭된 탬플릿 이미지와 연계되어 저장된 설명 내지는 서술에 대한 편집을 통하여 해당되는 전시 대상 작품에 관한 스타일 분석 큐레이션이 생성될 수도 있다.

【0099】본 발명의 일 실시형태에서, 상기 스타일 분석 모델은 CNN(convolution neural network) 계열의 네트워크를 포함할 수 있으며, 이미지(예를 들어, 전시 대상 작품의 이미지 또는 사전에 설정된 템플릿 이미지)를 입력으로

하고, 입력된 이미지로부터 형상 특징(윤곽과 같은 geometry, 칼라 톤 등)을 추출하여 각각의 클래스(class)에 해당되는 확률(confidence)을 산출하는 분류(classification) 태스크를 수행하는 CNN(convolution neural network) 네트워크의 적어도 일부를 포함할 수 있으며, 이러한 CNN 네트워크에서는 이미지가 입력되는 입력단으로부터 출력단을 향하여 레이어(layer)가 깊어짐에 따라, 각각의 레이어를 형성하는 채널(channel)의 개수가 증가하면서 각각의 채널의 행렬 차원(예를 들어, 화소 집합을 표현하는 행렬 차원)은 줄어들게 되며, 이때, CNN 네트워크의 특정한 레이어를 형성하는 서로 다른 채널을 형성하는 feature map(특징 맵, feature map)은 입력된 이미지의 서로 다른 특징(서로 다른 윤곽과 같은 geometry, 서로 다른 컬러 톤 등, feature 1~4)을 추출한 것으로 이해될 수 있으며, 본 발명의 일 실시형태에서는, 특정한 CNN 네트워크, 즉, 가중치가 고정된 CNN 네트워크의 특정한 레이어의 결과 값(예를 들어, 특정한 레이어를 형성하는 서로 다른 필터 또는 커널 내지는 서로 다른 필터 또는 서로 다른 커널이 적용된 합성 곱, 예를 들어, 각각의 채널, 필터, 커널은 각각 서로 다른 윤곽을 포함하는 geometry 또는 컬러 톤 등을 추출할 수 있음)을 형성하는 feature map이 표현하는 서로 다른 특징(feature 1~4) 사이의 상관관계(correlation)를 스타일(style)로 규정할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서는 서로 다른 특징(feature 1~4) 사이의 상관관계(correlation)를 표현한 Gram 매트릭스를 산출하고 산출된 Gram 매트릭스 자체를, 해당되는 이미지의 스타일 정보를 함축한 것으로 이해할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 전시 대상 작품을 형성하는 이미지와 사전에 설정된

템플릿 이미지는 동일한 파라메타 또는 가중치가 고정된 CNN 계열의 네트워크(예를 들어, VGG net)로 입력될 수 있으며, 이때, 상기 CNN 계열의 네트워크의 깊이 방향을 따라 특정된 레이어에서 서로 다른 특징(feature 1~4)을 추출하기 위한 서로 다른 필터 내지는 커널이 적용되면서, 각각 전시 대상 작품을 형성하는 이미지와 사전에 설정된 템플릿 이미지로부터 서로 다른 특징(feature 1~4)을 추출한 특징 맵(feature map)이 생성될 수 있으며, 이들 전시 대상 작품의 이미지와 템플릿 이미지로부터 추출된 특징(feature 1~4) 사이에서 각각의 특징 간의 상관관계(correlation, 특징 feature 1~4 간의 내적 dot product)가 표현된 Gram 매트릭스가 산출될 수 있다.

【0100】 본 발명의 일 실시형태에서, 상기 Gram 매트릭스(G_{ij})는 이하와 같이 표현될 수 있으며, 특정한 CNN 네트워크의 특정한 레이어 L을 형성하는 채널의 총 개수 k에 대해 서로 다른 특징 i와 j 사이의 상관관계를 표현하도록 각각의 특징 i와 j 사이의 내적(dot product)을 행렬의 원소로 하는 행렬의 형태로 표현될 수 있다.

【0101】

【0102】 예를 들어, 상기 Gram 매트릭스는 특정한 CNN 네트워크(예를 들어, CNN 네트워크의 일 유형으로서의 VGG net)의 특정한 레이어를 선정하고 특정한 레이어를 형성하는 서로 다른 채널 내지는 서로 다른 특징 i와 j의 내적으로 산출되

는 Gram 매트릭스를 산출하고, 각각의 전시 대상 작품의 이미지로부터 산출된 Gram 매트릭스와 템플릿 이미지로부터 산출된 Gram 매트릭스의 서로 다른 Gram 매트릭스의 각 원소 사이의 차분 값에 기반하여 전시 대상 작품의 이미지와 템플릿 이미지 사이의 유사도를 분석할 수 있으며, 보다 구체적으로, 전시 대상 작품의 이미지로부터 산출된 Gram 매트릭스와 템플릿 이미지로부터 산출된 Gram 매트릭스의 서로 다른 Gram 매트릭스의 각 원소 사이의 차분 값의 총합 또는 Gram 매트릭스의 각 원소 사이의 차분 값의 제곱의 총합 또는 Gram 매트릭스의 각 원 사이의 차분 값의 제곱의 총합을 scaling factor로 나누거나 또는 평균한 MSE(mean squared error, 평균제곱오차)를 산출함으로써, 전시 대상 작품의 이미지와 템플릿 이미지 사이의 상호 유사도 스코어 내지는 상호 매칭 적합도를 산출할 수 있으며, 전시 대상 작품의 이미지와 템플릿 이미지의 스타일 정보를 함축한 Gram 매트릭스의 각 원소 사이의 차분 값의 총합(또는 차분 값의 제곱의 총합 등)이 최소화되는 템플릿 이미지를 해당되는 전시 대상 작품의 이미지와 매칭되는 것으로 판단할 수 있으며, 이와 같이 전시 대상 작품의 이미지와 매칭되는 것으로 선정된 템플릿 이미지와 연계하여 저장된 설명 내지는 서술, 그러니까, 사전에 설정해둔 다수의 템플릿 이미지 각각과 연계하여 저장된 것으로, 다수의 템플릿 이미지 각각과 연계하여 저장된 각각의 템플릿 이미지의 스타일에 관한 설명 내지는 서술을 직접 해당되는 전시 대상 작품의 이미지에 관한 스타일 분석 큐레이션으로 생성하거나 또는 다수의 템플릿 이미지 각각과 연계하여 저장된 각각의 템플릿 이미지의 스타일에 관한 설명 내지는 서술에 관한 편집을 통하여 전시 대상 작품의 이미지에 관한 스타일 분석 큐레이션을

생성할 수도 있다.

【0103】본 발명의 일 실시형태에서, 상기 객체 분석 모델은, 전시 대상 작품을 형성하는 이미지 상에 등장하는 객체를 인식할 수 있으며, 객체 인식(object detection)을 구현할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 상기 객체 분석 모델은, 전시 대상 작품을 형성하는 이미지 상에 설정된 관심 영역(ROI, region of interest) 상에 객체의 존재 여부에 관한 예측 확률 및 관심 영역으로부터 상기 객체의 경계를 둘러싸는 바운딩 박스(bounding box, Bbox)의 정보를 산출하기 위한 신경망 네트워크를 포함할 수 있다. 예를 들어, 상기 객체 분석 모델은 Faster R-CNN 네트워크를 포함할 수 있으며, 보다 구체적으로, 상기 객체 분석 모델은 입력된 전시 대상 작품을 형성하는 이미지로부터 합성곱을 통하여 고차원의 이미지의 특성을 추출한 특성 벡터(feature vector) 또는 특성 맵(feature map)을 추출할 수 있으며, 특성 벡터(feature vector) 또는 특성 맵(feature map)은 고차원의 이미지 특성과 함께, 입력된 전시 대상 작품의 이미지의 위치 정보를 포함하고 있다는 점에서, 입력된 전시 대상 작품의 이미지로부터 추출된 특성 벡터(feature vector) 또는 특성 맵(feature map)은 객체가 존재할 가능성이 있는 관심 영역(ROI)을 탐색하기 위한 RPN(Region Proposal Network) 네트워크와, 각각의 관심 영역 내에 수확 대상의 농작물 객체가 존재하는지 여부에 관한 예측 확률을 산출하기 위한 Detection 네트워크에서 공유될 수 있다.

【0104】예를 들어, 상기 RPN 네트워크에서는 객체 분석 모델로 입력된 전시 대상 작품의 이미지로부터 추출된 특성 벡터(feature vector) 또는 특성 맵

(feature map) 상으로 서로 다른 형상 비율 및 스케일을 갖는 앵커 박스(anchor box)를 슬라이딩 윈도(sliding window) 방식으로 이동시키면서 합성곱을 수행하여 또 다른 이미지 특성(intermediate feature)을 추출한 특성 벡터(feature vector) 또는 특성 맵(feature map)을 생성하고, 각각의 앵커 박스(anchor box) 내에 객체가 존재하는지 여부에 관한 이진 분류(cls layer)과 앵커 박스(anchor box)로부터 해당되는 객체의 경계를 둘러싸는 보다 정확한 바운딩 박스(Bbox)의 위치 및 크기에 관한 정보를 추출하기 위한 리그레션(regression, reg layer)을 수행할 수 있고, 각각의 앵커 박스 내지는 바운딩 박스(Bbox)와 객체의 존재 여부에 관한 이진 분류의 정보는, 각각의 앵커 박스 내지는 바운딩 박스(Bbox) 내에 존재하는 객체에 관한 인식(detection) 내지는 분류(classification)에 관한 예측 확률을 산출하기 위한 Detection 네트워크로 입력될 수 있고, 상기 Detection 네트워크에서는 RPN 네트워크로부터 예측 결과를 참조하여 앵커 박스 내지는 바운딩 박스(Bbox)의 특성 벡터(feature vector) 내지는 특성 맵(feature map)을 전결합층에 입력하여 각각의 앵커 박스 내지는 바운딩 박스(Bbox) 내에 존재하는 객체에 대한 인식 내지는 분류에 관한 예측 확률을 산출할 수 있다(classifier). 이와 같이, 본 발명의 일 실시형태에 따른 객체 분석 모델은, 전시 대상 작품을 형성하는 이미지 상으로부터 포착된 객체의 인식(detection) 내지는 분류와 함께, 객체의 경계를 바운딩 박스(Bbox)로 예측하는 객체 인식(object detection)을 수행할 수 있으며, 예를 들어, 바운딩 박스(Bbox)의 중심 위치(P)와 바운딩 박스(Bbox)의 가로 x 세로의 크기에 관한 예측 값을 산출할 수 있다.

【0105】본 발명의 일 실시형태에서, 상기 객체 분석 모델이 입력된 전시 대상 작품을 형성하는 이미지 상으로부터 포착된 객체의 인식(detection) 내지는 분류(classification)에 관한 예측 확률을 추론하는데 그치지 않고, 더 나아가, 이미지 상에서 포착된 객체의 경계를 둘러싼 바운딩 박스(Bbox)의 위치 및 크기까지 추론하는 것은, 전시 대상 작품을 형성하는 이미지의 전체적인 분위기 내지는 감성에 영향을 줄 수 있는 전시 대상 작품의 이미지 상에 등장하는 다수의 객체들 사이의 상호 위치 관계 및 대소 관계를 고려하여 단순히 전시 대상 작품의 이미지 상에 등장하는 객체들 각각에 대한 인식 내지는 분류에 그치지 않고, 더 나아가, 전시 대상 작품을 형성하는 이미지의 전제적인 분위기 또는 감성에 관한 추가적인 정보를 추출하기 위한 것으로 이해할 수 있으며, 이런 점에서 본 발명의 일 실시형태에서, 상기 객체 분석 모델은, 전시 대상 작품을 형성하는 이미지 상에 등장하는 다수의 객체들 사이의 상대적인 위치 관계 및 대소 관계를 포함하는 다수의 객체들 사이의 공간 배치를 예측하는 것으로 이해될 수 있으며, 예를 들어, 전시 대상 작품의 중앙 위치에 인접하게 배치되는 객체일수록, 그리고, 상대적으로 넓은 영역을 점유하는 객체일수록, 해당되는 전시 대상 작품의 주제 또는 테마와 인접한 것으로 이해될 수 있으며, 예를 들어, 전시 대상 작품을 형성하는 이미지 상에서 주된 객체로 이해될 수 있고, 이와 달리, 전시 대상 작품의 중앙 위치로부터 멀리 이격되어 배치되는 객체일수록, 그리고, 상대적으로 좁은 영역을 점유하는 객체일수록, 해당되는 전시 대상 작품의 주제 또는 테마와는 거리가 있으며, 전시 대상 작품의 전체적인 분위기 내지는 감성에는 영향을 줄 수 있는 보조적인 객체(보조 객체)로 이해될

수 있다.

【0106】 예를 들어, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 상기 객체 분석 모델로부터 산출되는 공간 배치의 예측 결과로부터 전시 대상 작품의 중앙 위치에 인접하게 배치되는 객체일수록, 그리고, 상대적으로 넓은 영역을 점유하는 객체일수록, 해당되는 전시 대상 작품의 주제 또는 테마와 인접한 주된 객체로 추론할 수 있으며, 이와 달리, 전시 대상 작품의 중앙 위치로부터 멀리 떨어져 배치되는 객체일수록, 그리고, 상대적으로 좁은 영역을 점유하는 객체일수록, 해당되는 전시 대상 작품의 주제 또는 테마와는 먼 보조 객체로 추론할 수 있다.

【0107】 본 발명의 일 실시형태에서, 상기 의도 분석 모델은 전시 대상 작품을 형성하는 이미지 자체가 아닌, 전시 대상 작품에 부수되는 메타 데이터로서, 전시 대상 작품을 기획한 작가로부터 제공된 주관적인 의견으로서 작가의 작업 의도, 작가의 창작 의도, 작가의 철학 등이 포함된 작가 노트, 또는 작가 외의 전문가로부터의 작품 평, 전시 대상 작품을 기획한 작자(또는 작가 외의 전문가)가 명명한 작품 명을 포함하는 작품 캡션 정보(예를 들어, 작품 명, 크기, 매체 등)에 관한 메타 데이터로서 텍스트 데이터를 입력으로 하는 의도 분석 모델을 포함할 수 있다.

【0108】 예를 들어, 본 발명의 일 실시형태에서, 전시 대상 작품의 일 특징(feature)으로, 의도 분석 모델로부터 추출되어 전시 대상 작품에 관한 전체적인 큐레이션에 포함되는 의도 분석 큐레이션이란, 전시 대상 작품을 통하여 전시 대상 작품을 감상하는 전시 대상 작품의 소비자를 향하여 전달하고자 하는 주제, 테마,

분위기, 감성 등을 총괄적으로 아우르는 포괄적인 개념을 의미할 수 있으며, 예를 들어, 전시 대상 작품을 감상하는 소비자와의 상호 작용을 통하여 전달하고자 하는 주제, 테마, 분위기, 감성 등을 포괄적으로 표현한 개념으로 이해될 수 있다.

【0109】예를 들어, 본 발명의 일 실시형태에서, 상기 의도 분석 모델은, 전시 대상 작품을 형성하는 이미지 자체가 아닌, 이미지에 부수되는 텍스트 형태의 메타 데이터를 입력으로 하여, 해당되는 전시 대상 작품을 형성하는 이미지의 일 특징(feature)으로, 의도 분석에 관한 특징(feature)을 추출해내고, 추출된 특징으로부터 전시 대상 작품의 큐레이션의 일부를 형성하는 의도 분석 큐레이션을 생성할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, 상기 의도 분석 모델은 텍스트-투-텍스트의 자연어 처리를 구현하기 위한 언어 모델을 포함할 수 있다.

【0110】본 발명의 일 실시형태에서, 상기 의도 분석 모델은, 전시 대상 작품을 형성하는 이미지에 부수되는 텍스트 형태의 메타 데이터를 형성하는 다수의 문장 집합으로부터 입력된 메타 데이터를 함축한 요약 생성을 구현할 수 있으며, 예를 들어, 입력된 메타 데이터를 형성하는 다수의 문장 집합으로부터 중요한 문장을 추출하여 요약하는 추출 요약(extractive summarization) 이라기 보다는, 입력된 메타 데이터를 형성하는 다수의 문장 집합으로부터 중요한 문장을 추려 요약을 생성하는 추출 요약과는 달리, 입력된 메타 데이터를 형성하는 다수의 문장 집합을 의역(paraphrasing)하여 요약을 생성하는 생성 요약(abstractive summarization)을 구현할 수 있다.

【0111】상기 의도 분석 모델은 텍스트 형태로 입력된 메타 데이터의 전체적인 문맥을 포함하는 컨텍스트 표현을 추출하는 것으로 이해될 수 있으며, 예를 들어, RNN(Recurrent Neural Network)과 LSTM(Long Short-Term Memory) 네트워크와 같은 시퀀스-투-시퀀스(sequence to sequence) 아키텍처를 포함하거나 또는 인코더와 디코더를 포함하는 트랜스포머 아키텍처를 포함할 수 있다. 예를 들어, RNN(Recurrent Neural Network) 또는 LSTM 네트워크(Long Short Term Memory)와 같은 시퀀스-투-시퀀스 아키텍처를 이용하여 컨텍스트 벡터(context vector)를 추출할 수 있으며, 예를 들어, RNN 또는 LSTM 기반의 시퀀스-투-시퀀스(sequence to sequence) 네트워크를 이용하여, 입력된 메타 데이터를 형성하는 단어 시퀀스에 대해, 각각의 시간 스텝 마다 입력 임베딩을 순차적으로 입력하고, 최종적인 시간 스텝에서 출력되는 일정한 크기의 컨텍스트 벡터를 생성할 수 있다(예를 들어, 다 대 일 구조의 네트워크).

【0112】본 발명의 다양한 실시형태에서, 상기 의도 분석 모델은 forward language 모델인 GPT 또는 bi-directional language 모델인 BERT를 포함할 수 있다. 예를 들어, 상기 의도 분석 모델은 bi-directional language 모델인 BERT를 포함할 수 있으며, 상기 BERT에서 입력 임베딩은, 토큰 임베딩(token embedding), 세그먼트 임베딩(segmentation embedding), 위치 임베딩(positional embedding)을 포함할 수 있으며, 이러한 입력 임베딩을 입력으로 하여, 상기 BERT는 모든 토큰(단어)의 표현 값을 출력할 수 있다. 이때, 상기 입력 임베딩에서는 첫번째 문장의 시작 부분에 [CLS] 토큰을 추가하고, 문장의 끝 부분에 [SEP] 토큰을 추가할 수 있

으며, 이때, [CLS] 토큰에 대한 표현 값은 입력된 전시 대상 작품의 메타 데이터로서 텍스트 시퀀스에 대한 전체적인 문맥(context) 정보를 포함하는 컨텍스트 표현에 해당될 수 있다. 다시 말하면, 본 발명의 일 실시형태에서, 입력된 메타 데이터의 전체적인 문맥 정보를 포함하는 컨텍스트 표현의 생성을 위한 자연어 처리와 관련하여, BERT 모델에서는 임베딩된 입력 문장의 첫번째 토큰으로 [CLS] 토큰을 추가할 수 있으며, [CLS] 토큰에 관한 표현 값은 입력된 메타 데이터로서 문장 집합을 형성하는 각각의 문장 표현을 취합할 수 있으며, BERT 모델에서 [CLS] 토큰의 표현 값을 문장 전체에 관한 컨텍스트 표현으로 추출할 수 있으며, 입력된 메타 데이터에 관한 의도 분석의 특징(feature)을 추출한 컨텍스트 표현으로 출력할 수 있다.

【0113】 예를 들어, 상기 의도 분석 모델은 forward language 모델인 GPT를 포함할 수 있으며, 상기 GPT에서는 어텐션 블록(attention block)을 다수로 쌓아 올린 멀티 헤드 어텐션(multi-head attention)과 피드 포워드 네트워크(feedforwad network)를 포함하는 디코더 블록을 누적하여 다수로 쌓아 올리고, 각 디코더 블록의 결과 값이 다음으로 연결된 디코더 블록의 입력으로 들어가도록 다수의 디코더 블록이 연결될 수 있으며, 최후의 디코더 블록의 결과 값으로부터 최초의 디코더 블록으로 입력된 메타 데이터로서 메타 데이터를 형성하는 텍스트 또는 메타 데이터를 형성하는 텍스트에 관한 임베딩 표현에 대한 전체적인 문맥 정보를 포함하는 컨텍스트 표현(context vector)가 산출될 수 있다. 상기 디코더 블록은 멀티 헤드 어텐션(multi-head attention)과 피드 포워드 네트워크(feedforwad network)를 포

함할 수 있으며, 이외에 skip connection을 위한 Add와 임베딩 표현(또는 토큰)에 대한 정규화(예를 들어, 임베딩 표현 또는 토큰을 0~1 사이로 정규화)를 위한 Layer Normalization(Layer Norm)을 포함할 수 있다.

【0114】 예를 들어, 본 발명의 일 실시형태에 따른 의도 분석 모델로서 언어 모델에는 토큰 임베딩(token embedding), 위치 임베딩(positional embedding), 세그먼트 임베딩(segment embedding)이 합산된 입력 임베딩이 입력될 수 있으며, 상기 입력 임베딩은 문장 시퀀스의 시작에 해당되는 스페셜 토큰(예를 들어, Start token)으로부터 문장 시퀀스의 종료 이후의 스페셜 토큰(예를 들어, Extract token)을 포함할 수 있으며, 최후의 디코더 블록의 Extract 토큰의 표현 값을 문장 전체에 관한 컨텍스트 표현으로 추출할 수 있으며, 입력된 메타 데이터에 관한 의도 분석의 특징(feature)을 추출한 컨텍스트 표현으로 출력할 수 있다.

【0115】 이와 같이, 본 발명의 일 실시형태에서, 상기 의도 분석 모델은 전시 대상 작품을 형성하는 이미지 자체가 아닌, 전시 대상 작품을 형성하는 이미지에 부수되는 텍스트 형태의 메타 데이터를 형성하는 다수의 문장 집합으로부터 입력된 메타 데이터를 함축한 요약 생성을 구현하는 것으로 이해될 수 있으며, 이런 의미에서 상기 전시 대상 작품의 일 특징(feature)으로 의도 분석에 관한 특징(feature) 또는 이에 기반한 의도 분석 큐레이션이란 전시 대상 작품을 관람하는 전시 대상 작품의 소비자를 향하여 전달하고자 하는 주제, 테마, 분위기, 감성 등을 총괄적으로 아우르는 협의의 의도 분석에 관한 특징을 포함하고, 보다 넓은 광의의 개념에서는 전시 대상 작품을 형성하는 이미지 자체가 아닌, 전시 대상 작품

의 이미지에 부수되는 텍스트 형태의 메타 데이터에 관하여 생성된 요약의 개념으로 이해될 수 있으며, 예를 들어, 전시 대상 작품을 통하여 전시 대상 작품의 소비자에게 전달하고자 하는 의도의 개념에서 벗어나, 전시 대상 작품의 이미지에 부수되는 메타 데이터의 요약 생성으로 이해될 수 있다.

【0116】 앞서 설명된 감성 분석에서는 전시 대상 작품을 형성하는 이미지 자체를 입력으로 하는 이미지 기반의 감성 분석과 이미지 자체가 아닌, 이미지에 부수되는 메타 데이터로서 텍스트를 입력으로 하는 텍스트 기반의 감성 분석을 구현하면서, 이들 이미지 기반의 감성 분석의 예측과 텍스트 기반의 감성 분서의 예측을 취합하여, 전시 대상 작품의 감성 분석에 관한 특징 내지는 전시 대상 작품의 감성 분석에 관한 특징에 기반한 전시 대상 작품의 감성 분석에 관한 감성 분석 큐레이션이 생성될 수 있다.

【0117】 본 발명의 다양한 실시형태에서, 전시 대상 작품으로부터 감성 분석에 관한 특징(feature)을 추출하기 위한 감성 분석은, 전시 대상 작품을 형성하는 이미지 자체를 입력으로 하는 이미지 기반의 감성 분석만으로 구현하면서 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터를 입력으로 하는 텍스트 기반의 감성 분석은 구현하지 않되, 텍스트 기반의 감성 분석은, 전시 대상 작품을 형성하는 이미지에 수반되는 메타 데이터로서 텍스트 데이터를 입력으로 하여, 텍스트-투-텍스트를 구현하는 의도 분석을 통하여 구현될 수 있으며, 예를 들어, 본 발명의 일 실시형태에서, 상기 의도 분석을 통하여 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터로부터 요약 생성을 구현하면서 의도 분석의 결과로서

전시 대상 작품의 일 특징(feature)으로, 전시 대상 작품을 통하여 전시 대상 작품의 소비자를 향하여 전달하고자 하는 감성에 관한 특징(feature) 내지는 감성 분석 큐레이션이 생성될 수 있고, 예를 들어, 전시 대상 작품으로부터 추출된 다른 특징들, 색감 분석, 스타일 분석, 객체 분석의 특징들과는 달리, 메타 데이터로서 텍스트 데이터로 포함된 감성 분석의 특징이, 감성 분석과 의도 분석에서 이중으로 추출되면서 전시 대상 작품의 큐레이션에서 다른 특징들에 비하여 불규형적으로 부각 내지는 강조되는 바이어스(bias)를 회피하기 위하여, 본 발명의 일 실시형태에서는 메타 데이터로서 텍스트 데이터에 포함된 감성과 관련된 특징을, 감성 분석과 의도 분석을 통하여 이중으로 추출하지 않고, 예를 들어, 감성 분석에서는 이미지 기반의 감성 분석만으로 감성 분석의 특징을 추출하고 의도 분석에서는 텍스트 기반의 감성 분석만으로 감성 분석의 특징을 추출함으로써, 후술하는 바와 같이, 각각의 서로 다른 모델로부터 추출된 서로 다른 분석의 특징 내지는 서로 다른 분석의 특징에 기반하는 서로 다른 분석의 큐레이션을 취합하면서, 감성 분석과 의도 분석을 통하여 이중으로 추출된 메타 데이터에 포함된 텍스트 형태의 감성 분석의 특징이 과도하게 부각 내지는 강조되는 편향성이나 바이어스를 회피하고, 서로 다른 모델로부터 추출된 서로 다른 특징들이 균형적으로 최종적인 전시 대상 작품의 큐레이션에 포함되도록 할 수 있다.

【0118】본 발명의 다양한 실시형태에서, 입출력 관계에 관하여 이미지-투-텍스트(멀티-모달, multi-modal)를 구현하는 이미지 기반의 감성 분석 보다는, 입출력 관계에 관하여 텍스트-투-텍스트(또는 요약 생성, 유니-모달, uni-modal)를

구현하는 텍스트 기반의 감성 분석에서 예측 정확도가 보다 높을 수 있다는 점을 고려하고, 또한, 후술하는 바와 같이, 본 발명의 일 실시형태에서, 전시 대상 작품의 큐레이션은 전시 공간의 추천을 포함할 수 있으며, 이와 같이 전시 공간의 추천에서는 전시 대상 작품의 감성 분석의 특징이 주요한 추천 근거를 형성할 수 있다는 점에서, 감성 분석의 특징 내지는 이에 근거하는 감성 분석의 큐레이션에 보다 많은 가중치(예를 들어, 감성 분석의 예측에 대해 다른 분석의 예측 보다 상대적으로 높은 가중치를 부여함)를 부여할 수 있다는 점과, 이미지-투-텍스트(멀티-모달)를 구현하는 이미지 기반의 감성 분석 보다 상대적으로 높은 예측 정확도가 예상되는 텍스트-투-텍스트(요약 생성, 유니-모달)를 구현하는 텍스트 기반의 감성 분석의 예측에 상대적으로 높은 가중치를 부여한다는 점에서, 본 발명의 일 실시형태에서는, 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트 기반의 감성 분류가 실질적으로 이중으로 구현되도록 할 수 있고, 예를 들어, 메타 데이터를 입력으로 하여 메타 데이터에 관한 요약 생성을 추출하기 위한 실질적으로 동일한 신경망 네트워크 또는 신경망 아키텍처를 포함하는 감성 분석 모델 및 의도 분석 모델을 통하여 이중으로 추출된 메타 데이터에 포함된 감성 분석의 특징(feature)이 상대적으로 높은 가중치로서 전시 대상 작품의 큐레이션에 포함되도록 할 수 있다. 이와 같은 실시형태에서, 상기 감성 분석 모델(예를 들어, 텍스트 기반의 감성 분석 모델)과 의도 분석 모델은, 전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터를 입력으로 하여 메타 데이터에 관한 요약 생성을 구현하기 위한 실질적으로 동일 유사한 신경망 네트워크 또는 신경망 아키텍처를 포함할 수 있으며, 예를

들어, 상기 감성 분석 모델과 의도 분석 모델은, 입력된 메타 데이터에 대한 요약 생성을 위한 네트워크를 공유할 수 있으며, 다만, 상기 감성 분석 모델(예를 들어, 텍스트 기반의 감성 분석 모델)은, 추출 대상이 되는 감성 분석의 특징과 관련된 단어 표현 내지는 임베딩 표현에 보다 높은 가중치를 부여하면서 감성 분석과는 무관한 특징과 관련된 단어 표현 내지는 임베딩 표현에는 보다 낮은 가중치 또는 실질적으로 가중치를 부여하지 않는 필터링-아웃(filtering-out)을 통하여 감성 분석의 특징을 추출할 수 있도록 학습된 가중치 세트(감성 분석의 특징에 편향되도록 학습된 가중치 세트)를 포함할 수 있으며, 이와 달리, 상기 의도 분석 모델은 감성 분석에 치우치지 않고 입력된 메타 데이터를 형성하는 문장 집합 또는 단어 집합에서 전체적인 문맥을 포괄하는 요약을 생성할 수 있다. 예를 들어, 본 발명의 일 실시형태에서, 이미지 기반의 감성 분석 모델과 텍스트 기반의 감성 분석 모델은 각각의 모델로부터의 예측을 취합하기 위한 concatenate layer로부터 개시되는 네트워크를 통하여 이들 서로 다른 이미지 기반의 감성 분석 모델과 텍스트 기반의 감성 분석 모델의 예측을 취합할 수 있고, 이때, 예를 들어, 동일한 감성 분류의 공간 상으로 매핑된 이미지 기반의 감성 분석 모델의 예측과 텍스트 기반의 감성 분석 모델의 예측 사이의 유클리드 거리에 따라 이들 예측이 취합되는 것으로 이해될 수 있으며, 예를 들어, 감성 분류의 공간 상에 매핑되는 임베딩 표현에서 감성 외의 다른 특징(feature)과 관련된 표현은 대체로 가중치가 낮게 부여되거나 또는 실질적으로 가중치가 부여되지 않고 필터링-아웃(filtering-out)될 수도 있다.

【0119】다시 말하면, 본 발명의 일 실시형태에서, 상기 의도 분석 모델은, 적어도 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 감성에 관한 특징을 추출하고, 상기 감성 분석 모델은, 전시 대상 작품의 이미지 데이터를 입력으로 하는 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하는 이미지 기반의 감성 분석 모델은 포함하되,

【0120】전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델은 포함하지 않을 수 있다.

【0121】본 발명의 다른 실시형태에서, 상기 의도 분석 모델은, 적어도 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 감성에 관한 특징을 추출하고,

【0122】상기 감성 분석 모델은, 전시 대상 작품의 이미지 데이터를 입력으로 하는 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하는 이미지 기반의 감성 분석 모델과 함께,

【0123】전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델을 포함하되,

【0124】상기 의도 분석 모델과 텍스트 기반의 감성 분석 모델은, 입력된 메타 데이터를 함축한 요약 생성을 위한 네트워크를 공유할 수 있다. 이때, 상기 텍

스트 기반의 감성 분석 모델은 입력된 메타 데이터를 함축한 요약 생성에 대해, 추출 대상이 되는 감성 분석의 특징과 관련된 임베딩 표현에 상대적으로 높은 가중치를 부여하면서 감성 분석의 특징과 무관한 임베딩 표현에 상대적으로 낮은 가중치를 부여하거나 또는 가중치를 부여하지 않으면서 필터링-아웃(filtering-out)시키도록 학습된 가중치 세트를 포함할 수 있다.

【0125】앞서 설명된 바와 같이, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은 각각 서로 다른 특징을 추출하기 위한 서로 다른 분석 모델을 포함할 수 있으며, 예를 들어, 입출력 관계에 관하여 이미지-투-텍스트(멀티-모달, multi-modal)를 구현하기 위한 색감 분석 모델, 입출력 관계에 관하여 이미지-투-텍스트(멀티-모달, multi-modal)를 구현하기 위한 이미지 기반의 감성 분석과 선택적으로 텍스트-투-텍스트(유니-모달, uni-modal)를 구현하기 위한 텍스트 기반의 감성 분석을 포함하는 감성 분석 모델, 입출력 관계에 관하여 이미지-투-텍스트(멀티-모달, multi-modal)를 구현하기 위한 스타일 분석 모델, 입출력 관계에 관하여 이미지-투-텍스트(멀티-모달, multi-modal)를 구현하기 위한 객체 분석 모델, 그리고, 입출력 관계에 관하여 텍스트-투-텍스트(유니-모달, uni-modal)를 구현하기 위한 의도 분석 모델을 포함할 수 있으며, 이들 서로 다른 모델로부터의 예측 결과 또는 예측 결과에 기반한 분석 큐레이션을 취합하기 위한 concatenate layer로부터 개시되는 네트워크를 포함할 수 있으며, 앞서 설명된 바와 같이, 서로 다른 모델로부터의 예측 결과는 공통적으로 텍스트 임베딩 표현(서로 다른 분석의 특징 또는 서로 다른 분석의 큐레이션에 관한 텍스트 임베딩 표현,

동일한 텍스트 임베딩 공간 상으로 매핑되는 텍스트 임베딩 표현)을 포함할 수 있으며, 이에 따라, 서로 다른 모델로부터 출력되는 텍스트 임베딩 표현(서로 다른 분석의 특징 또는 서로 다른 분석의 큐레이션에 관한 텍스트 임베딩 표현, 동일한 텍스트 임베딩 공간 상으로 매핑되는 텍스트 임베딩 표현)을 취합하기 위한 언어 모델을 포함할 수 있으며, 예를 들어, 상기 concatenate layer로부터 개시되는 언어 모델은, 레이블이 부여되지 않은 unlabeled 데이터 세트(unlabeled large corpus, 웹 텍스트와 같은 대규모 텍스트)를 학습 데이터로 하여 학습된 범용적인 자연어 처리에 대한 사전 지식을 습득한 pre-trained 모델을 포함할 수 있으며, 각각의 서로 다른 모델로부터 출력되는 텍스트 임베딩 표현을 취합하고 범용적인 자연어 처리에 관한 사전 지식으로부터 이들 서로 다른 모델로부터 출력되는 텍스트 임베딩 표현으로부터 해당되는 전시 대상 작품에 관한 큐레이션을 생성할 수 있다. 예를 들어, 본 발명의 일 실시형태에서 상기 서로 다른 분석 모델로부터의 예측을 취합하기 위한 언어 모델은, 서로 다른 모델로부터의 출력을 concatenate 및 flatten 시키기 위한 layer, linear fully connected layer, 임베딩 표현 내지는 토큰의 정규화를 위한 normalization layer 등을 포함할 수 있다.

【0126】 예를 들어, 본 발명의 일 실시형태에서, 상기 전시 대상 작품에 관하여 생성되는 큐레이션은, 전시 대상 작품에 관한 서로 다른 분석의 결과를 취합하여 생성된 전시 대상 작품에 관한 설명 또는 기술(예를 들어, 서로 다른 분석의 특징에 기반한 협의의 큐레이션)을 포함할 수 있으며, 이와 함께, 서로 다른 분석의 결과를 취합하여 생성된 전시 대상 작품에 관한 설명 또는 기술(예를 들어, 서

로 다른 분석의 특징에 기반한 협의의 큐레이션)을 입력하여 하여, 해당되는 전시 대상 작품의 전시 공간에 관한 추천(예를 들어, 서로 다른 분석의 특징에 기반한 협의의 큐레이션과 함께 전시 공간의 추천을 포함하는 광의의 큐레이션)을 포함할 수 있다. 달리 표현하면, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 전시 대상 작품의 서로 다른 분석의 특징을 추출하기 위한 서로 다른 분석 모델로서, 앞서 설명된 바와 같은, 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델, 객체 분석 모델, 및 의도 분석 모델을 포함할 수 있으며, 이들 서로 다른 분석 모델로부터의 예측 결과를 취합하여 생성된 전시 대상 작품에 관한 설명, 서술 또는 협의의 큐레이션을 입력으로 하여 전시 대상 공간을 추천하기 위한 공간 추천 모델을 더 포함할 수 있다.

【0127】예를 들어, 상기 공간 추천 모델은 앞서 설명된 다수의 분석 모델로부터의 예측 결과에 기반한 전시 대상 작품에 관한 설명, 서술 또는 협의의 큐레이션과 같은 텍스트 데이터를 입력으로 하여 공간 추천에 관하여 사전에 설정된 클래스 중에서 매칭 적합도가 높은 클래스를 추천해줄 수 있으며, 예를 들어, 사전에 설정된 클래스로서 카페, 병원, 사무실, 휴게 공간, 주거 공간과 같은 대분류의 카테고리와, 특정된 대분류의 카테고리에 속하는 소분류의 카테고리로서 예를 들어, 주거 공간(대분류의 카테고리)의 거실, 아이방, 침실방과 같은 소분류의 카테고리를 포함하거나 또는 대분류의 카테고리로부터 소분류의 카테고리로 계층적인 hierarchical 구조를 취할 수도 있다.

【0128】본 발명의 일 실시형태에서, 상기 공간 추천 모델은, 전시 대상 작품에 관하여 서로 다른 분석 모델로부터 추출된 특징(feature) 내지는 특징(feature)에 기반한 분석 큐레이션에 관한 텍스트 데이터를 입력으로 하여 사전에 설정된 다수의 클래스와의 매칭 적합도에 관한 확률을 예측할 수 있고, 가장 높은 예측 확률이 산출된 클래스에 대해 전시 대상 작품과 매칭되는 것으로 판단하고, 가장 높은 예측 확률이 산출된 클래스에 관한 공간을 전시 대상 작품이 전시될 최적의 공간으로 추천해줄 수 있다.

【0129】예를 들어, 상기 공간 추천 모델은, 서로 다른 분석 모델로부터 추출된 특징(feature) 내지는 특징(feature)에 기반한 분석 큐레이션에 관한 텍스트 데이터를 입력으로 하여 사전에 설정된 다수의 클래스 중에서 매칭되는 클래스를 추천해주는 다중 분류를 구현할 수 있으며, 예를 들어, 상기 공간 추천 모델은 다중 분류의 classifier(분류기)를 포함할 수 있다. 예를 들어, 본 발명의 일 실시형태에서 상기 공간 추천 모델은 서로 다른 분석 모델로부터 추출된 특징(feature) 내지는 특징(feature)에 기반한 분석 큐레이션에 관한 텍스트 시퀀스 또는 단어 집합 또는 문장 집합(이들의 임베딩 표현)을 입력으로 하여 차원 축소된 임베딩 표현을 추출하기 위한 텍스트 인코더를 포함할 수 있으며, 예를 들어, 상기 공간 추천 모델은 입력된 텍스트 시퀀스 또는 단어 집합 또는 문장 집합을 함축하거나 또는 이들 텍스트 시퀀스 또는 단어 집합 또는 문장 집합의 전체적인 문맥(context)을 포함하도록 차원 축소된 임베딩 표현을 추출할 수 있으며, 이러한 컨텍스트 표현이 매핑된 다차원의 텍스트 임베딩 공간 상에서 각각의 클래스에 속하는 공간에 관한

임베딩 표현과의 유클리드 거리에 기반하여 해당되는 전시 대상 작품에 대한 추천을 구현하는 것으로 이해될 수 있으며, 이때, 각각의 클래스에 속하는 공간에 관한 임베딩 표현과 동일한 텍스트 임베딩 공간 상으로 매핑되는 컨텍스트 표현 사이의 상관 관계 내지는 이들 각각의 클래스에 속하는 공간에 관한 임베딩 표현과 서로 다른 분석 모델로부터 추출된 특징(feature) 내지는 특징(feature)에 기반한 분석 큐레이션에 관한 텍스트 시퀀스 또는 단어 집합 또는 문장 집합으로부터 추출된 컨텍스트 표현 사이의 상관 관계를 형성하는 텍스트 임베딩 공간은, 상기 공간 추천 모델의 학습을 통하여 학습될 수 있다.

【0130】도 19에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, random noise z를 입력으로 하여 합성 데이터(생성 대상인 큐레이션)를 생성하기 위한 Generator G와 Generator G로부터 생성된 합성 데이터와 원본 데이터를 입력으로 하여 합성 데이터(Fake)와 원본 데이터(Real) 사이를 판별하기 위한 Discriminator D를 포함하는 GAN(Generative Adversarial Network)을 설명하기 위한 도면이며, 각각의 손실 함수(G loss, D loss)로부터 산출된 Gradient descent의 역전파(backpropagation) 알고리즘을 통한 Generator G와 Discriminator D의 학습을 설명하기 위한 도면이 도시되어 있다.

【0131】도 20에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반

의 생성형 AI 네트워크를 설명하기 위한 도면으로, random noise z를 입력으로 하여 합성 데이터(생성 대상인 큐레이션)를 생성하기 위한 Generator G와 Generator G로부터 생성된 합성 데이터와 원본 데이터를 입력으로 하여 합성 데이터(Fake)와 원본 데이터(Real) 사이를 판별하기 위한 Discriminator D를 포함하는 GAN(Generative Adversarial Network)을 설명하기 위한 도면이 도시되어 있다.

【0132】도 21에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Time series based GAN에 적용될 수 있는 시퀀스(sequence) 모델의 예시로서, Time series based GAN에 포함될 수 있는 네트워크로서 RNN(Recurrent Neural Network)과 LSTM(Long Short Term Memory)의 구조를 예시적으로 보여주는 도면이 도시되어 있다.

【0133】도 22에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Time series based GAN에 적용될 수 있는 시퀀스(sequence) 모델의 예시로서, LSTM의 연쇄를 포함하는 Time series based GAN(Sequence GAN)의 Generator G의 예시를 보여주는 도면이 도시되어 있다.

【0134】도 23에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반

의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Generator G와 Discriminator D가 시퀀스 모델로서 LSTM(Long Short Term Memory)을 포함하거나, 또는 Generator G가 시퀀스 모델로서 LSTM(Long Short Term Memory)을 포함하고 Discriminator D가 CNN(Convolution Neural Network)을 포함하는 Sequence GAN의 아키텍처를 설명하기 위한 도면과, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Generator G가 MC search(Monte Carlo Search Tree)를 통한 보상(Reward)으로부터 정책 또는 정책 함수를 업-데이트 시키는 강화 학습 또는 강화 학습의 policy gradient를 설명하기 위한 도면이 도시되어 있다.

【0135】도 24에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Generator G가 latent vector z와 이전 시간 스텝의 출력으로부터 각각의 시간 스텝의 출력으로서 시계열의 합성 데이터(생성 대상인 큐레이션)를 생성하도록 LSTM(Long Short Term Memory)의 시퀀스 모델을 포함하는 구성을 설명하기 위한 도면이 도시되어 있다.

【0136】도 25에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반

의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN을 형성하는 Discriminator D가 Generator G로부터 출력되는 시계열의 합성 데이터(생성 대상인 큐레이션)를 입력으로 하여, 각각의 시간 스텝의 합성 데이터(생성 대상인 큐레이션)를 입력으로 하는 LSTM(Long Short Term Memory)의 시퀀스 모델의 결과에 대한 vote로부터 최종적인 판별을 추론하는 것을 설명하기 위한 도면이 도시되어 있다.

【0137】 도 26에는 본 발명의 큐레이션 생성 시스템을 구현하기 위한 다른 실시형태로서, 큐레이션의 생성을 위한 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 설명하기 위한 도면으로, 시퀀스 데이터 또는 시계열 데이터(생성 대상인 큐레이션)의 처리를 위한 Sequence GAN의 Discriminator D를 형성하는 CNN(Convolution Neural Network)을 설명하기 위한 도면이 도시되어 있다.

【0138】 도 27에는 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템의 입력으로서 전시 대상 작품을 형성하는 이미지 자체가 아닌, 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 예시적으로 보여주는 도면이 도시되어 있다.

【0139】 도 28에는 도 27에 예시된 메타 데이터로서 텍스트 데이터와 함께, 전시 대상 작품을 형성하는 이미지 데이터를 입력으로 하여, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템으로부터 생성된 큐레이션(전시 공간의 추천을 포함하는 광의의 큐레이션)을 예시적으로 보여주는 도면이 도시되어 있다.

【0140】본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은, 전시 대상 작품에 대한 서로 다른 특징 내지는 서로 다른 특징에 기반한 분석 큐레이션을 추출하기 위한 서로 다른 분석 모델을 포함할 수 있으며, 본 발명의 다양한 실시형태에 따른 큐레이션 생성 시스템은 생성형 AI 네트워크로서, GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크를 포함할 수 있다.

【0141】예를 들어, GAN(Generative Adversarial Network)은 기본적으로 2개의 neural network, Generator G와 Discriminator D로 구성될 수 있으며, Generator G는 random 노이즈(또는 latent vector) z를 입력으로 하여 training 데이터 분포와 유사한 합성 데이터(synthetic data)를 생성할 수 있으며, 이때, Discriminator D는 입력된 데이터가 real(원본 데이터) 인지 fake(합성 데이터, 가상 데이터) 인지를 구분할 수 있으며, Generator G의 목표는 Discriminator D가 real로 판단할 만큼 real에 가까운 합성 데이터를 생성하는 것이고, Discriminator D의 목표는 Generator G에서 생성된 합성 데이터를 분별하는 것으로, GAN은 하기와 같은 목표 함수를 통해 학습을 진행할 수 있으며, Generator G와 Discriminator D가 목표 함수의 minmax를 달성하도록 학습될 수 있으며, D(x)는 입력 데이터 x에 대해 Discriminator D가 real로 판단할 확률을 의미할 수 있다.

【0142】$\min(G)\max(D)\ V(G, D) = E_{x \sim pdata(x)}[\log D(x)] + E_{z \sim pz(z)}[\log(1-D(G(z)))]$

【0143】이와 같이, 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은 GAN(Generative Adversarial Network) 기반의 생성형 AI 네트워크로서 입력된 단어

시퀀스의 다음 단어를 예측하면서 시간 스텝의 전진에 따라 다음에 등장할 단어(다음 단어)를 순차적으로 예측하면서, 예를 들어, 이전 시간 스텝에서 나온 모든 단어들로부터 다음에 나올 다음 단어를 예측하는 auto-regressive한 방식으로 다음 단어를 추론하면서 전시 대상 작품의 큐레이션을 시간 스텝의 전진에 따라 순차적으로 생성할 수 있으며, 전시 대상 작품의 큐레이션을 시계열적인 데이터로 하는 Time series based GAN을 포함할 수 있다.

【0144】예를 들어, 상기 Time series based GAN으로서 Sequence GAN(SeqGAN)에서는 시계열적인 시퀀스 데이터(sequential data)를 입력으로 하는 시퀀스 아키텍처(시퀀스 모델)를 포함할 수 있으며, 예를 들어, RNN(Recurrent Neural Network)이나 LSTM(Long Short Term memory)과 같이 매 시간 스텝 마다 순차적으로 입력되는 시계열적인 시퀀스 데이터를 입력으로 하여, 각각의 시간 스텝 마다 출력(예를 들어, 다음 단어에 관한 예측)을 산출할 수 있다. 보다 구체적으로, 상기 Sequence GAN(SeqGAN)은 GAN(Generative Adversarial Network) 기반으로 원본 데이터의 분포를 추종하는 합성 데이터로서 시계열 데이터를 생성하기 위한 AI 네트워크로서, random 노이즈(또는 latent vector) z를 입력으로 하여 training data set와 유사한 합성 데이터를 생성하도록 학습되는 Generator G와, 원본 데이터 또는 Generator G로부터 생성된 합성 데이터를 입력으로 하여, 원본 데이터 또는 합성 데이터를 판별하도록 학습되는 Discriminator D를 포함할 수 있으며, 예를 들어, 본 발명의 일 실시형태에서 상기 Sequence GAN을 형성하는 Generator G와 Discriminator D 중에서 적어도 어느 하나의 neural network는 앞서

설명된 바와 같은 시퀀스 아키텍처(또는 시퀀스 모델)을 포함할 수 있으며, 예를 들어, RNN(Recurrent Neural Network)이나 LSTM(Long Short Term memory)과 같은 시퀀스 아키텍처(또는 시퀀스 모델)을 포함할 수 있다.

【0145】 예를 들어, 본 발명의 일 실시형태에서 상기 Sequence GAN을 형성하는 Generator G와 Discriminator D 중에서, Generator G는 LSTM(Long Short Term Memory)을 포함할 수 있으며, 매 시간 스텝 마다 random 노이즈(또는 latent vector) z를 입력으로 하여 각각의 시간 스텝에서의 출력(다음 단어에 관한 예측)으로부터 시계열적인 시퀀스 데이터(sequential data, 예를 들어, 전시 대상 작품의 큐레이션)를 합성 데이터로서 생성할 수 있으며, 상기 Sequence GAN을 형성하는 Generator G와 Discriminator D 중에서, Discriminator D는 LSTM(Long Short Term Memory)을 포함할 수 있으며, Generator G로부터 생성된 시계열적인 시퀀스 데이터(sequential data)를 형성하는 매 시간 스텝에서의 value를 입력으로 하여, 각각의 시간 스텝에서의 출력을 산출할 수 있고, 이때, 상기 Discriminator D는 각각의 시간 스텝에서의 출력들에 대한 vote를 통하여 최종적인 출력으로서 real과 fake의 판단을 출력할 수 있다.

【0146】 본 발명의 일 실시형태에서, 합성 데이터로서 시계열 데이터(예를 들어, 전시 대상 작품의 큐레이션)를 생성하기 위한 Sequence GAN은 앞서 설명된 바와 같이 시퀀스 아키텍처(예를 들어, LSTM)를 포함하는 Generator G를 포함할 수 있으며, 이때 Generator G는 강화 학습(RL, Reinforcement Learning)을 통하여 학습될 수 있고, 상기 Generator G의 강화 학습(RL, Reinforcement Learning)에서는

매 시간 스텝에서 다음 시간 스텝의 출력으로 예측된 액션(action, next action, 예를 들어, 다음 단어에 관한 예측)에 대해 액션(action, 예를 들어, 다음 단어에 관한 예측)의 결과로서의 보상(reward) 내지는 액션의 결과로서 환경(Environment)으로부터 주어지는 보상(reward)를 Discriminator D로부터 제공받을 수 있다. 다만, 본 발명의 일 실시형태에서, 상기 Discriminator D는 Generator G로부터 다음 시간 스텝에서의 출력으로 예측된 액션(action, 다음 단어에 관한 예측)에 대해 매 시간 스텝 마다 보상(reward)을 제공한다기 보다는 하나의 에피소드(episode)가 종료된 시점에서 다음 시간 스텝에서의 출력으로 예측된 액션(action)에 대한 보상(reward)을 제공할 수 있으며, 이때, Generator G로부터 예측된 다음 스텝에서의 출력으로서의 액션에 대한 즉각적인 보상(immediate reward)이 제공되지 않고, 처음 시간 스텝으로부터 최종적인 마지막 시간 스텝 까지 하나의 에피소드(episode)가 종료된 이후에 비로소 Generator G로부터 예측된 다음 스텝에서의 출력으로서의 액션에 대한 보상이 제공될 수 있다. 이와 같이 각각의 시간 스텝에서의 즉각적인 보상이 이루어지지 않고 하나의 에피소드가 종료된 이후에 비로소 보상이 제공되는 Generator G의 학습에서는 Generator G로부터 예측된 다음 시간 스텝에서의 출력으로서의 액션(action) 이후로부터 마지막 시간 스텝까지, 그러니까, 하나의 에피소드의 종료까지 Generator G의 출력은 MC Search Tree(Monte Carlo Search Tree)로부터 예측될 수 있으며, 이와 같이 Generator G로부터 출력된 다음 시간 스텝의 액션 이후로의 액션은 Generator G로부터 출력된다기 보다는 MC Search Tree(Monte Carlo Search Tree)로부터 예측될 수 있다.

【0147】본 발명의 일 실시형태에서, 전시 대상 작품의 큐레이션의 생성을 위한 sequence GAN에서는 매 시간 스텝 마다 이전 시간 스텝에서 출력된 모든 단어(단어 임베딩, 토큰 임베딩) 또는 모든 표현으로부터 다음에 등장할 단어 또는 표현을 예측할 수 있으며, 이때, 합성 데이터의 생성을 위한 Generator G와 입력된 데이터에 대한 real(원본 데이터)/fake(합성 데이터)의 판별을 위한 Discriminator D의 서로 다른 두 개의 네트워크를 포함하는 GAN 기반의 생성형 AI 네트워크이면서, 강화 학습(RL, Reinforcement Learning)을 통하여 학습되는 Generator G를 포함하여, 시계열적인 데이터로서, 전시 대상 작품의 큐레이션을 생성하기 위한 sequence GAN에서는 매 시간 스텝 마다 예측된 다음 단어 또는 다음 표현에 대한 예측에 대해 매 시간 스텝 마다 즉각적인 보상이 제공되지 않을 수 있으며, 예를 들어, 하나의 에피소드가 종료된 이후에 비로소 Generator G로부터 예측된 다음 스텝에서의 출력으로서 액션(예를 들어, 다음 단어 또는 다음 표현에 관한 예측)에 대한 보상이 제공될 수 있으며, 예를 들어, 본 발명의 일 실시형태에서 전시 대상 작품의 큐레이션 전체의 예측이 종료된 이후에 또는 전시 대상 작품의 큐레이션을 형성하는 열 단위 또는 문장 단위의 예측이 종료된 이후에 비로소 Generator G로부터 예측된 다음 스텝에서의 출력으로서 액션(예를 들어, 다음 단어 또는 다음 표현에 관한 예측)에 대한 보상이 제공될 수 있다. 그리고, 이와 같이 각각의 시간 스텝에서의 즉각적인 보상이 이루어지지 않고 하나의 에피소드가 종료된 이후(예를 들어, 전시 대상 작품의 큐레이션 전체의 예측이 종료된 이후 또는 전시 대상 작품의 큐레이션을 형성하는 열 단위 또는 문장 단위의 예측이 종료된

이후)에 비로소 보상이 제공되면서 지연된 보상이 제공되는 Generator G의 학습에서는 Generator G로부터 예측된 다음 시간 스텝에서의 출력으로서의 액션(action, 예를 들어, 다음 단어 또는 다음 표현에 관한 예측) 이후로부터 마지막 시간 스텝까지, 그러니까, 하나의 에피소드의 종료(예를 들어, 전시 대상 작품의 큐레이션 전체의 예측 종료 또는 전시 대상 작품의 큐레이션을 형성하는 열 단위 또는 문장 단위의 예측 종료)까지 Generator G의 출력은 MC Search Tree(Monte Carlo Search Tree)로부터 예측될 수 있으며, 이와 같이 Generator G로부터 출력된 다음 시간 스텝의 액션 이후로의 액션(다음 단어 또는 다음 표현에 관한 예측)은 Generator G로부터 출력된다기 보다는 MC Search Tree(Monte Carlo Search Tree)로부터 예측될 수 있다.

【0148】상기 Generator G의 강화 학습(RL, Reinforcement Learning)에서는 에이전트(agent)로서 Generator G로부터 예측된 특정 액션(예를 들어, 전시 대상 작품의 큐레이션을 형성할 다음 단어 또는 다음 표현)에 대해 Discriminator D로부터 상대적으로 많은 보상이나 또는 양의 보상이 제공되면, 향후에 상태(이전 시점에서 예측된 Generator G의 출력, 예를 들어, Generator G의 출력으로, 이전 시간 스텝에서 예측된 모든 단어) 관찰을 통하여 동일 내지는 유사한 상태에서 같은 액션이 취해질 확률을 높이는 방향으로 정책 함수(Policy) 내지는 정책 함수를 정의하는 파라메타를 갱신할 수 있으며, 이와 달리, 에이전트로서 Generator G로부터 취해진 특정 액션에 대해 Discriminator D로부터 상대적으로 적은 보상이나 또는 음의 보상이 제공되면, 향후에 상태(이전 시점에서 예측된 Generator G의 출력, 예

를 들어, Generator G의 출력으로, 이전 시간 스텝에서 예측된 모든 단어) 관찰로부터 동일 내지는 유사한 상태에서 같은 액션이 취해지는 확률을 낮추는 방향으로 정책 함수 내지는 정책 함수를 정의하는 가중치(파라메타)를 갱신할 수 있다. 이와 같이, 본 발명의 일 실시형태에서 Generator G의 강화 학습에서 목적 함수에 대한 그래디언트 어센트(Gradient Ascent) 또는 그래디언트 디센트(Gradient Descent)로부터 정책 함수(Policy)를 업-데이트 시킬 수 있다.

【0149】 본 발명의 일 실시형태에서, 합성 데이터 또는 가상 데이터로서 시계열 데이터를 생성하기 위한 Sequence GAN은 앞서 설명된 바와 같이 시퀀스 아키텍처(예를 들어, LSTM)를 포함하는 Generator G를 포함할 수 있으며, 이때, 상기 Discriminator D는 도에 도시된 바와 같이, LSTM(Long Short Term Memory)을 포함할 수 있으며, Generator G로부터 생성된 시계열적인 시퀀스 데이터(sequential data)를 형성하는 매 시간 스텝에서의 value를 입력으로 하여, 각각의 시간 스텝에서의 출력을 산출할 수 있고, 이때, 상기 Discriminator D는 각각의 시간 스텝에서의 출력들에 대한 vote를 통하여 최종적인 출력으로서 real과 fake의 판단을 출력할 수 있고, 다른 실시형태에서, 상기 Discriminator D는 CNN(Convolution Neural Network)을 포함할 수 있으며, CNN(Convolution Neural Network)에서 학습된 파라메타를 포함하는 필터(가중치 행렬, 커널)의 합성곱을 통하여 행열의 2차원 크기가 축소되면서 feature map이 추출될 수 있으며, feature map으로부터 Concat(concatenation)과 Multi-layer Perceptron을 통하여 Generator G로부터 생성된 시계열적 데이터로서 가상 데이터(또는 합성 데이터)에 대한 real/fake의 판

별이 출력될 수 있다.

【0150】 본 발명의 일 실시형태에 따른 큐레이션 생성 시스템은,

【0151】 생성 대상인 전시 대상 작품의 큐레이션을 시간 스텝의 전진에 따라 다음에 등장할 표현을 예측하면서 순차적으로 생성되는 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터로 생성하기 위한 Time series based GAN을 포함할 수 있다.

【0152】 예를 들어, 상기 Time series based GAN으로서 Sequence GAN을 형성하는 Generator G와 Discriminator D 중에서 적어도 어느 하나의 네트워크는, RNN(Recurrent Neural Network) 또는 LSTM(Long Short Term memory)의 시퀀스 아키텍처를 포함할 수 있다.

【0153】 예를 들어, 상기 Generator G는 시간 스텝의 전진에 따라 각각의 시간 스텝 마다 random 노이즈 또는 latent vector를 입력으로 하여 각각의 시간 스텝에서 다음에 등장할 표현에 관한 예측으로부터 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터로서 전시 대상 작품의 큐레이션을 순차적으로 생성할 수 있다.

【0154】 예를 들어, 상기 Discriminator D는 Generator G로부터 시간 스텝의 전진에 따라 각각의 시간 스텝 마다 생성된 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터를 입력으로 하여, 각각의 시간 스텝에서의 출력을 산출하되, 각각의 시간 스텝에서의 출력들에 대한 vote를 통하여 최종적인 출력으로서 real과 fake의 판단을 출력할 수 있다.

【0155】예를 들어, 상기 Generator G는 강화 학습(RL, Reinforcement Learning)의 에이전트(agent)로서,

【0156】상기 Discriminator D로부터 상대적으로 많은 보상(reward)이나 또는 양(positive)의 보상이 제공되면 향후의 상태 관찰(state observation)로서, 이전 시간 스텝에서 예측된 모든 표현에 관한 관찰을 통하여 동일 유사한 상태에서 동일한 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현으로 동일한 표현을 출력할 확률을 높이는 방향으로 정책 함수(policy function)를 갱신하며,

【0157】상기 Discriminator D로부터 상대적으로 적은 보상이나 또는 음(negative)의 보상이 제공되면 향후의 상태 관찰(state observation)로서, 이전 시간 스텝에서 예측된 모든 표현에 관한 관찰을 통하여 동일 유사한 상태에서 동일한 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현으로 동일한 표현을 출력할 확률을 낮추는 방향으로 정책 함수(policy function)를 갱신할 수 있다.

【0158】예를 들어, 상기 Discriminator D는 시간 스텝의 전진에 따라 매 시간 스텝 마다 상기 Generator G로부터 취해진 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현에 대한 예측에 대해 즉각적인 보상(immediate reward)을 제공하지 않고,

【0159】하나의 에피소드(Episode)가 종료된 이후로서, 전시 대상 작품의 큐레이션을 형성하는 열 단위 또는 문장 단위의 예측이 종료된 이후에 비로소 상기 Generator G로부터 취해진 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음

표현에 대한 지연된 보상을 제공할 수 있다.

【0160】 예를 들어, 상기 Generator G의 학습에서, 상기 Generator G로부터 취해진 액션으로서, 전시 대상 작품의 큐레이션을 형성할 다음 표현에 관한 현재 시간 스텝에서의 예측 이후로 다음 시간 스텝에서의 예측으로부터 하나의 에피소드(Episode)의 종료까지의 Generator G로부터 취해지는 다음 시간 스텝 이후의 예측은 MC Search Tree(Monte Carlo Search Tree)로부터 예측될 수 있다.

【0161】 예를 들어, 상기 Discriminator D는 시간 스텝의 전진에 따라 상기 Generator G로부터 취해진 액션으로서, 전시 대상 작품의 큐레이션을 형성할 다음 표현에 관한 현재 시간 스텝에서의 예측과, 전시 대상 작품의 큐레이션을 형성할 다음 시간 스텝 이후로 하나의 에피소드(Episode)의 종료까지 MC Search Tree(Monte Carlo Search Tree)로부터의 예측을 취합한 에피소드(Episode) 단위로 보상(reward)을 제공할 수 있다.

【0162】 본 발명의 일 실시형태에서, 상기 시계열 데이터는 앞서 설명된 바와 같은 Sequence GAN(SeqGAN)과 같은 다양한 Time series based GAN을 통하여 생성될 수 있다.

【0163】 본 발명은 첨부된 도면에 도시된 실시예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 수 있을 것이다.

【부호의 설명】

【0164】Multi-modal: 분석 모델의 입출력 데이터가 서로 다른 유형의 이미지 데이터/텍스트 데이터로 구성된 멀티-모달

Uni-modal: 분석 모델의 입출력 데이터가 서로 같은 유형의 이미지 데이터/텍스트 데이터로 구성된 유니-모달

convolution: 컨볼루션 레이어

pooling: 풀링 레이어

Gram matrix: 그램 매트릭스

RPN: Region Proposal Network

context vector: 컨텍스트 벡터 또는 컨텍스트 표현

【청구범위】

【청구항 1】

전시 대상 작품을 형성하는 이미지 데이터 및 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터로부터 전시 대상 작품의 서로 다른 특징들을 추출하여 텍스트 형태의 설명 또는 서술을 포함하는 큐레이션을 생성하기 위한 큐레이션 생성 시스템.

【청구항 2】

제1항에 있어서,

전시 대상 작품을 형성하는 이미지에 부수되는 메타 데이터로서 텍스트 데이터는,

전시 대상 작품을 기획한 작가의 작업 의도, 작가의 창작 의도, 또는 작가의 철학이 포함된 작가 노트에 관한 텍스트 데이터; 또는

전시 대상 작품에 대한 전문가의 작품 평, 작가 또는 전문가가 명명한 작품명, 전시 대상 작품의 사이즈, 또는 전시 대상 작품을 형성하는 재료에 관한 매체 정보를 포함하는 작품 캡션 정보에 관한 텍스트 데이터를 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 3】

제1항에 있어서,

전시 대상 작품의 서로 다른 특징들을 추출하기 위한 다수의 분석 모델; 및

상기 다수의 분석 모델로부터 추출된 서로 다른 특징들을 취합하여 상기 큐레이션을 생성하기 위한 신경망 네트워크를 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 4】

제3항에 있어서,

상기 다수의 분석 모델은, 상기 전시 대상 작품을 형성하는 이미지 상에 표현된 서로 다른 특징들을 추출하기 위한 것으로,

색감 또는 컬러 톤에 관한 특징을 추출하기 위한 색감 분석 모델;

감성에 관한 특징을 추출하기 위한 감성 분석 모델;

스타일에 관한 특징을 추출하기 위한 스타일 분석 모델;

전시 대상 작품 상에 표현된 객체(object)에 관한 특징을 추출하기 위한 객체 분석 모델; 및

전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 주제, 테마, 분위기 또는 감성을 포괄하는 의도에 관한 특징을 추출하기 위한 의도 분석 모델;

을 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 5】

제4항에 있어서,

상기 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델, 객체 분석 모델 및 의도 분석 모델은, 전시 대상 작품을 형성하는 이미지 데이터 또는 전시 대상

작품을 형성하는 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여, 동일한 텍스트 임베딩 공간 상으로 매핑되는 텍스트 임베딩 표현을 출력하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 6】

제4항에 있어서,

상기 다수의 분석 모델 각각에 대한 입출력 관계에 관하여,

상기 색감 분석 모델, 감성 분석 모델, 스타일 분석 모델 및 객체 분석 모델은, 전시 대상 작품의 이미지 데이터를 입력으로 하여 서로 다른 특징에 관한 텍스트 임베딩 표현을 출력하면서 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하고,

상기 의도 분석 모델은, 전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여 입력된 메타 데이터의 요약 생성을 통하여 텍스트 임베딩 표현을 출력하면서 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 7】

제6항에 있어서,

상기 감성 분석 모델은, 상기 감성 분석 모델에 대한 입출력 관계에 관하여,

전시 대상 작품의 이미지 데이터를 입력으로 하여 감성 분석의 예측에 관한 텍스트 임베딩 표현을 출력하면서 이미지-투-텍스트의 멀티-모달(multi-modal)을

구현하는 이미지 기반의 감성 분석 모델을 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 8】

제7항에 있어서,

상기 감성 분석 모델은, 상기 감성 분석 모델에 대한 입출력 관계에 관하여,

전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하여 입력된 메타 데이터로부터 감성 분석의 예측에 관한 텍스트 임베딩 표현을 출력하면서 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델을 더 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 9】

제8항에 있어서,

상기 감성 분석 모델은, 상기 이미지 기반의 감성 분석 모델의 예측 결과와, 상기 텍스트 기반의 감성 분석 모델의 예측 결과를 취합하여 전시 대상 작품으로부터 추출된 감성에 관한 특징으로 출력하기 위한 신경망 네트워크를 더 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 10】

제4항에 있어서,

상기 의도 분석 모델은, 상기 메타 데이터로서 텍스트 시퀀스, 단어 집합 또는 문장 집합으로부터 입력된 메타 데이터를 함축한 요약 생성을 구현하는 것을 특

징으로 하는, 큐레이션 생성 시스템.

【청구항 11】

제4항에 있어서,

상기 의도 분석 모델은, 적어도 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 감성에 관한 특징을 추출하고,

상기 감성 분석 모델은,

전시 대상 작품의 이미지 데이터를 입력으로 하는 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하는 이미지 기반의 감성 분석 모델은 포함하되,

전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력으로 하는 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델은 포함하지 않는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 12】

제4항에 있어서,

상기 의도 분석 모델은, 적어도 전시 대상 작품을 감상하는 소비자를 향하여 전달하고자 하는 감성에 관한 특징을 추출하고,

상기 감성 분석 모델은,

전시 대상 작품의 이미지 데이터를 입력으로 하는 이미지-투-텍스트의 멀티-모달(multi-modal)을 구현하는 이미지 기반의 감성 분석 모델과 함께,

전시 대상 작품의 이미지에 부수되는 메타 데이터로서 텍스트 데이터를 입력

으로 하는 텍스트-투-텍스트의 유니-모달(uni-modal)을 구현하는 텍스트 기반의 감성 분석 모델을 포함하되,

상기 의도 분석 모델과 텍스트 기반의 감성 분석 모델은, 입력된 메타 데이터를 함축한 요약 생성을 위한 네트워크를 공유하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 13】

제12항에 있어서,

상기 텍스트 기반의 감성 분석 모델은 입력된 메타 데이터를 함축한 요약 생성에 대해, 추출 대상이 되는 감성 분석의 특징과 관련된 임베딩 표현에 상대적으로 높은 가중치를 부여하면서 감성 분석의 특징과 무관한 임베딩 표현에 상대적으로 낮은 가중치를 부여하거나 또는 가중치를 부여하지 않으면서 필터링-아웃(filtering-out)시키도록 학습된 가중치 세트를 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 14】

제4항에 있어서,

상기 색감 분석 모델은,

전시 대상 작품을 형성하는 이미지로부터 색감 또는 컬러 톤(color tone) 정보를 표현하기 위한 히스토그램(histogram) 또는 히스토그램 정보에 기반하여 전시 대상 작품을 형성하는 전체 이미지 상에서 표현되는 색감 또는 컬러 톤 정보를 추

출하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 15】

제14항에 있어서,

상기 색감 분석 모델은,

전시 대상 작품의 이미지를 형성하도록 서로에 대해 합성되는 3채널 이미지(R,G,B 3채널 이미지 또는 Y,Cb,Cr 3채널 이미지) 각각에 대해 화소 값 별로 등장하는 화소 개수를 표현하는 제1 내지 제3 히스토그램 또는 제1 내지 제3 히스토그램 정보에 기반하여 전시 대상 작품의 색감 또는 컬러 톤에 관한 특징을 추출하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 16】

제4항에 있어서,

상기 스타일 분석 모델은,

전시 대상 작품을 형성하는 이미지로부터 추출된 스타일 정보와, 사전에 설정된 다수의 템플릿 이미지 각각으로부터 추출된 스타일 정보 사이의 유사도 분석에 기반하여, 전시 대상 작품의 스타일에 관한 특징을 추출하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 17】

제16항에 있어서,

상기 스타일 분석 모델은,

전시 대상 작품을 형성하는 이미지와 가장 높은 유사도 스코어가 산출된 템플릿 이미지를 전시 대상 작품과 매칭되는 템플릿 이미지로 하여, 매칭된 템플릿 이미지와 연계된 스타일에 관한 특징을 탐색하여 탐색된 결과에 따라 전시 대상 작품의 스타일에 관한 특징으로 출력하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 18】

제4항에 있어서,

상기 객체 분석 모델은,

전시 대상 작품의 이미지 상에 등장하는 객체에 대한 인식 또는 분류와 함께, 전시 대상 작품의 이미지 상에 등장하는 다수의 객체들 사이의 상대적인 위치 관계 및 대소 관계를 포함하여 다수의 객체들 사이의 공간 배치를 예측하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 19】

제18항에 있어서,

상기 객체 분석 모델로부터 산출되는 공간 배치의 예측 결과로부터,

전시 대상 작품의 이미지 상에서 중앙 위치에 인접하게 배치되는 객체일수록, 또는 상대적으로 넓은 영역을 점유하는 객체일수록, 전시 대상 작품의 주체 또는 테마와 인접한 주된 객체로 추론하며,

전시 대상 작품의 이미지 상에서 중앙 위치로부터 멀리 떨어진 객체일수록,

또는 상대적으로 좁은 영역을 점유하는 객체일수록, 전시 대상 작품의 주체 또는 테마로부터 먼 보조 객체로 추론하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 20】

제1항에 있어서,

생성 대상인 전시 대상 작품의 큐레이션을 시간 스텝의 전진에 따라 다음에 등장할 표현을 예측하면서 순차적으로 생성되는 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터로 생성하기 위한 Time series based GAN을 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 21】

제20항에 있어서,

상기 Time series based GAN으로서 Sequence GAN을 형성하는 Generator G와 Discriminator D 중에서 적어도 어느 하나의 네트워크는, RNN(Recurrent Neural Network) 또는 LSTM(Long Short Term memory)의 시퀀스 아키텍처를 포함하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 22】

제21항에 있어서,

상기 Generator G는 시간 스텝의 전진에 따라 각각의 시간 스텝 마다 random 노이즈 또는 latent vector를 입력으로 하여 각각의 시간 스텝에서 다음에 등장할 표현에 관한 예측으로부터 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터로서

전시 대상 작품의 큐레이션을 순차적으로 생성하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 23】

제21항에 있어서,

상기 Discriminator D는 Generator G로부터 시간 스텝의 전진에 따라 각각의 시간 스텝 마다 생성된 시퀀스의 합성 데이터 또는 시계열적인 합성 데이터를 입력으로 하여, 각각의 시간 스텝에서의 출력을 산출하되, 각각의 시간 스텝에서의 출력들에 대한 vote를 통하여 최종적인 출력으로서 real과 fake의 판단을 출력하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 24】

제21항에 있어서,

상기 Generator G는 강화 학습(RL, Reinforcement Learning)의 에이전트(agent)로서,

상기 Discriminator D로부터 상대적으로 많은 보상(reward)이나 또는 양(positive)의 보상이 제공되면 향후의 상태 관찰(state observation)로서, 이전 시간 스텝에서 예측된 모든 표현에 관한 관찰을 통하여 동일 유사한 상태에서 동일한 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현으로 동일한 표현을 출력할 확률을 높이는 방향으로 정책 함수(policy function)를 갱신하며,

상기 Discriminator D로부터 상대적으로 적은 보상이나 또는 음(negative)의

보상이 제공되면 향후의 상태 관찰(state observation)로서, 이전 시간 스텝에서 예측된 모든 표현에 관한 관찰을 통하여 동일 유사한 상태에서 동일한 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현으로 동일한 표현을 출력할 확률을 낮추는 방향으로 정책 함수(policy function)를 갱신하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 25】

제24항에 있어서,

상기 Discriminator D는 시간 스텝의 전진에 따라 매 시간 스텝 마다 상기 Generator G로부터 취해진 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현에 대한 예측에 대해 즉각적인 보상(immediate reward)을 제공하지 않고,

하나의 에피소드(Episode)가 종료된 이후로서, 전시 대상 작품의 큐레이션을 형성하는 열 단위 또는 문장 단위의 예측이 종료된 이후에 비로소 상기 Generator G로부터 취해진 액션으로서 전시 대상 작품의 큐레이션을 형성할 다음 표현에 대한 지연된 보상을 제공하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 26】

제25항에 있어서,

상기 Generator G의 학습에서, 상기 Generator G로부터 취해진 액션으로서, 전시 대상 작품의 큐레이션을 형성할 다음 표현에 관한 현재 시간 스텝에서의 예측 이후로 다음 시간 스텝에서의 예측으로부터 하나의 에피소드(Episode)의 종료까지

의 Generator G로부터 취해지는 다음 시간 스텝 이후의 예측은 MC Search Tree(Monte Carlo Search Tree)로부터 예측되는 것을 특징으로 하는, 큐레이션 생성 시스템.

【청구항 27】

제26항에 있어서,

상기 Discriminator D는 시간 스텝의 전진에 따라 상기 Generator G로부터 취해진 액션으로서, 전시 대상 작품의 큐레이션을 형성할 다음 표현에 관한 현재 시간 스텝에서의 예측과, 전시 대상 작품의 큐레이션을 형성할 다음 시간 스텝 이후로 하나의 에피소드(Episode)의 종료까지 MC Search Tree(Monte Carlo Search Tree)로부터의 예측을 취합한 에피소드(Episode) 단위로 보상(reward)을 제공하는 것을 특징으로 하는, 큐레이션 생성 시스템.

【요약서】

【요약】

본 발명에서는 큐레이션 생성 시스템이 개시된다. 본 발명에 의하면, 전시 대상 작품을 형성하는 이미지 및 전시 대상 작품의 이미지에 부수되는 메타 데이터로서, 전시 대상 작품을 기획한 작가의 작업 의도, 작가의 창작 의도, 또는 작가의 철학이 포함된 작가 노트에 관한 메타 데이터 및/또는 전시 대상 작품에 대한 전문가의 작품 평, 작가 또는 전문가가 명명한 작품 명, 전시 대상 작품의 사이즈, 또는 전시 대상 작품을 형성하는 재료에 관한 매체 정보를 포함하는 작품 캡션 정보에 관한 메타 데이터로부터 전시 대상 작품의 서로 다른 특징들을 추출하여 텍스트 형태의 설명 또는 서술을 포함하는 협의의 큐레이션을 생성할 수 있으며, 또한 이와 같이 생성된 협의의 큐레이션을 입력으로 하여 전시 대상 작품의 전시 공간에 관한 추천을 포함하는 광의의 큐레이션을 생성할 수 있는 큐레이션 생성 시스템이 제공된다.

【대표도】

도 1a

【도면】

【도 1a】

【도 1b】

【도 2】

【도 3】

【도 4】

【도 5】

【도 6】

【도 7】

【도 8】

【도 9】

【도 10】

【도 11】

【도 12】

【도 13】

【도 14】

【도 15】

【도 16】

【도 17】

【도 18】

【도 19】

【도 20】

【도 21】

【도 22】

【도 23】

【도 24】

【도 25】

【도 26】

【도 27】

【도 28】