(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년02월13일
(11) 등록번호 10-2636717
(24) 등록일자 2024년02월07일

(51) 국제특허분류(Int. Cl.)
G06T 3/40 (2024.01) *G06N 3/08* (2023.01)
G06V 20/00 (2023.01)
(52) CPC특허분류
G06T 3/4046 (2024.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2022-0151336
(22) 출원일자 2022년11월14일
심사청구일자 2022년11월14일
(56) 선행기술조사문헌
Seung-Jin Ryu et al., "Estimation of linear transformation by analyzing the periodicity of interpolation", Pattern Recognition Letters, (2014.01.15.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
한림대학교 산학협력단
강원도 춘천시 한림대학길 1, 한림대학교(옥천동)
(72) 발명자
허종욱
강원도 춘천시 한림대학길 1 공학관 A1301호
최민기
강원도 춘천시 교동길 66-10 2층
(뒷면에 계속)
(74) 대리인
특허법인본

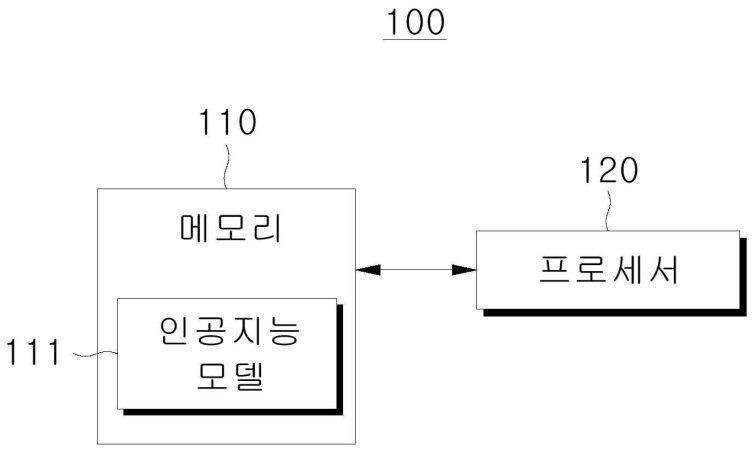
전체 청구항 수 : 총 9 항 심사관 : 이정은

(54) 발명의 명칭 **이미지의 변환을 추정하는 트랜스포머 기반의 전자 장치, 및 그 제어 방법**

(57) 요 약

전자 장치가 개시된다. 본 전자 장치는, 이미지에 적용된 변환을 추정하도록 훈련된 인공지능 모델을 포함하는 메모리, 타겟 이미지를 인공지능 모델에 입력하여, 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하는, 프로세서를 포함한다. 인공지능 모델은, 타겟 이미지를 주파수 영역으로 변환하여 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어, FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder)를 포함한다.

대 표 도 - 도2b



(52) CPC특허분류
G06T 3/4084 (2024.01)
G06V 20/95 (2022.01)
G06T 2207/20081 (2013.01)
G06T 2207/20084 (2013.01)
(72) 발명자
이상영
강원도 춘천시 방송길 70(온의동, 온의 롯데캐슬 스카이클래스) 107동 2002호

정희선
강원도 강릉시 입암로 29번길 11 101동 403호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711169460
과제번호	2020R1C1C1013433
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	3D 프린팅-스캐닝 콘텐츠 환경을 위한 멀티미디어 포렌식 원천기술 연구
기 여 율	1/1
과제수행기관명	한림대학교 산학협력
연구기간	2022.03.01 ~ 2023.02.28

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

전자 장치에 있어서,

이미지에 적용된 변환을 추정하도록 훈련된 인공지능 모델을 포함하는 메모리; 및

타겟 이미지를 상기 인공지능 모델에 입력하여, 상기 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하는, 프로세서;를 포함하고,

상기 인공지능 모델은,

상기 타겟 이미지를 주파수 영역으로 변환하여 상기 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어; 및

상기 FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder);를 포함하고,

상기 인공지능 모델은,

적어도 하나의 미분 필터 및 컨볼루션 레이어를 포함하는 전처리(preprocessing) 레이어;를 더 포함하고,

상기 프로세서는,

상기 타겟 이미지를 상기 전처리 레이어에 입력하여 제1 이미지 데이터를 획득하고,

상기 제1 이미지 데이터를 상기 FPE 레이어에 입력하여 제2 이미지 데이터를 획득하고,

상기 제2 이미지 데이터를 상기 트랜스포머 인코더에 입력하여 제3 이미지 데이터를 획득하는, 전자 장치.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 인공지능 모델은,

GELU 활성화 함수로 구성된 MLP(Multi-layer Perceptron)를 포함하는 회귀(regression) 레이어;를 포함하고,

상기 프로세서는,

상기 제3 이미지 데이터를 상기 회귀 레이어에 입력하고, 상기 회귀 레이어의 출력을 바탕으로 변환 행렬을 추정하는, 전자 장치.

청구항 4

제3항에 있어서,

상기 회귀 레이어는,

상기 추정된 변환 행렬과 정답에 해당하는 변환 행렬 간의 MSE(Mean Square Error)를 바탕으로 구성된 손실 함수에 따라 훈련되는, 전자 장치.

청구항 5

제1항에 있어서,

상기 FPE(Frequency Positional Encoding) 레이어는,

상기 타겟 이미지에 대하여 1차 이상의 미분이 수행된 잔차(u'_h)에 대하여 2D FFT(Fast Fourier Transform)를 수행하여 주파수 영역으로 변환된 이미지에 대한 함수 값인 $U_h(u, v)$를 획득하고,

상기 $U_h(u, v)$의 크기 성분을 일정 크기의 패치($x \in R^{\times 2}$)로 나누어 이하 수학식에 따른 시퀀스 z_0를 출력하고,

$$E \in \mathbb{R}^{p^2 \times d},\ E_{pos} \in \mathbb{R}^{n \times d}$$

$$Z_0 = [x_p^1 E; x_p^2 E; \ldots; x_p^n E] + E_{pos}$$

상기 n은 패치의 결과 수이고, 상기 p는 패치의 크기이고, 상기 E_{pos}는 위치 인코딩에 해당하는, 전자 장치.

청구항 6

제5항에 있어서,

상기 트랜스포머 인코더는,

상기 시퀀스 z_0에 대하여 어텐션 연산을 수행하여 Q(Query), K(Key), 및 V(Value)를 획득하고,

인코딩된 입력 스펙트럼이 있는 Q와 스펙트럼 도메인에 나타날 것으로 예상되는 피크 정보가 있는 K 간의 내적을 통해 확률 값을 산출하고,

상기 확률 값을 V에 투영하여 추정과 관련된 데이터를 획득하는, 전자 장치.

청구항 7

제1항에 있어서,

상기 트랜스포머 인코더는,

MSA(Multi-head Self-Attention) 레이어 및 MLP(Multi-Layer Perceptron) 레이어를 포함하는, 전자 장치.

청구항 8

이미지에 적용된 변환을 추정하도록 훈련된 인공지능 모델을 포함하는 전자 장치의 제어 방법에 있어서,

상기 전자 장치가, 타겟 이미지를 상기 인공지능 모델에 입력하는 단계; 및

상기 전자 장치가, 상기 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하는 단계;를 포함하고,

상기 인공지능 모델은,

상기 타겟 이미지를 주파수 영역으로 변환하여 상기 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어; 및

상기 FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder);를 포함하고,

상기 인공지능 모델은,

적어도 하나의 미분 필터 및 컨볼루션 레이어를 포함하는 전처리(preprocessing) 레이어;를 더 포함하고,

상기 전자 장치가, 상기 타겟 이미지를 상기 전처리 레이어에 입력하여 제1 이미지 데이터를 획득하고,

상기 전자 장치가, 상기 제1 이미지 데이터를 상기 FPE 레이어에 입력하여 제2 이미지 데이터를 획득하고,

상기 전자 장치가, 상기 제2 이미지 데이터를 상기 트랜스포머 인코더에 입력하여 제3 이미지 데이터를 획득하는, 전자 장치의 제어 방법.

청구항 9

전자 장치의 프로세서에 의해 실행되어 상기 전자 장치로 하여금 제8항의 제어 방법을 수행하도록 하는 컴퓨터 프로그램이 저장된 비일시적 컴퓨터 판독 가능 매체.

청구항 10

전자 장치에 있어서,

이미지에 적용된 변환을 추정하도록 훈련된 인공지능 모델을 포함하는 메모리; 및

타겟 이미지를 상기 인공지능 모델에 입력하여, 상기 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하는, 프로세서;를 포함하고,

상기 인공지능 모델은,

적어도 하나의 미분 필터 및 컨볼루션 레이어를 포함하는 전처리(preprocessing) 레이어;

상기 타겟 이미지를 주파수 영역으로 변환하여 상기 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어;

상기 FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder); 및

GELU 활성화 함수로 구성된 MLP(Multi-layer Perceptron)를 포함하는 회귀(regression) 레이어;를 포함하는, 전자 장치.

발명의 설명

기 술 분 야

[0001] 본 개시는 이미지의 변환을 추정하는 전자 장치에 관한 것으로, 보다 상세하게는, 트랜스포머 인코더를 포함하는 인공지능 모델을 바탕으로 변환 행렬을 추정하는 전자 장치에 관한 것이다.

배 경 기 술

[0002] 디지털 이미지는 전문가 외 대부분의 일반 사용자도 쉽게 편집 및 수정이 가능하다. 이러한 시각적 정보를 갖는 위조된 이미지는 대중매체, 소셜 미디어, 법적 증거로서 악의적으로 사용될 수도 있다. 이러한 이유로, 디지털 이미지의 조작 기록을 추적하기 위한 디지털 이미지 포렌식이 꾸준히 연구되어 왔다.

[0003] 디지털 이미지 포렌식은 디지털 이미지에 적용된 변환 작업이 결과 이미지에 몇 가지 독특한 흔적을 남긴다는 가정을 기반으로 연구된다. 디지털 영상의 기하학적 변환은 hole과 overlap 등의 문제를 발생시키기 때문에, nearest neighbor, bilinear, bicubic 등의 신호 보간법이 불가피하게 사용된다. 이러한 보간(interpolation)은 이미지에 고유한 흔적을 남기며, resampling을 추적하는 디지털 이미지 포렌식에서 중요한 단서로 작용한다. 그중 스케일링 및 회전과 같은 이미지 기하학적 변환은 이미지 처리 기법 중 가장 일반적으로 사용되며 디지털 이미지 포렌식에서 필수적인 문제이다.

[0004] 기하학적 변환을 추정하는 주요 방법론은 두 가지가 존재하며, 먼저 이미지의 2차 미분 신호를 기반으로 한 handcrafted 기반 기술이 제시되었다. 이는 고속 푸리에 변환(FFT)을 이용하여 보간법의 주기적 특성을 파악한 후 주파수 영역 분석을 이용하여 스케일링 및 회전 계수를 추정하였다.

[0005] 또 다른 기술은 CNN(Convolutional Neural Network)를 기반으로 전처리 계층을 갖는 프레임워크를 제안하고, 스케일링 계수를 추정하기 위해 분류 문제로 공식화하였다. 그러나 기하학적 변환은 본질적으로 연속적인 공간에서 정의되기 때문에 분류 기반 방법은 구별되는 조합의 수를 무한히 확장할 수 없다는 한계가 존재한다.

발명의 내용

해결하려는 과제

[0006] 본 개시는 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하여, 이미지의 변환/위조 여부, 변환 정도를 추정할 수 있는 전자 장치 및 그 제어 방법을 제공한다.

[0007] 본 개시는 이미지의 기하학적 변환 매개변수를 예측할 수 있는 엔드 투 엔드 트랜스포머 기반 추정기(estimator)를 포함하는 전자 장치 및 그 제어 방법을 제공한다.

[0008] 본 개시의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 개시의 다른 목적 및 장점들

은 하기의 설명에 의해서 이해될 수 있고, 본 개시의 실시 예에 의해 보다 분명하게 이해될 것이다. 또한, 본 개시의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0009] 본 개시의 일 실시 예에 따른 전자 장치는, 이미지에 적용된 변환을 추정하도록 훈련된 인공지능 모델을 포함하는 메모리, 타겟 이미지를 상기 인공지능 모델에 입력하여, 상기 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하는 프로세서를 포함한다. 상기 인공지능 모델은, 상기 타겟 이미지를 주파수 영역으로 변환하여 상기 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어, 상기 FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder)를 포함한다.

[0010] 상기 인공지능 모델은, 적어도 하나의 미분 필터 및 컨볼루션 레이어를 포함하는 전처리(preprocessing) 레이어를 더 포함할 수 있다. 이때, 상기 프로세서는, 상기 타겟 이미지를 상기 전처리 레이어에 입력하여 제1 이미지 데이터를 획득하고, 상기 제1 이미지 데이터를 상기 FPE 레이어에 입력하여 제2 이미지 데이터를 획득하고, 상기 제2 이미지 데이터를 상기 트랜스포머 인코더에 입력하여 제3 이미지 데이터를 획득할 수 있다.

[0011] 또한, 상기 인공지능 모델은, GELU 활성화 함수로 구성된 MLP(Multi-layer Perceptron)를 포함하는 회귀(regression) 레이어를 더 포함할 수 있다. 이때, 상기 프로세서는, 상기 제3 이미지 데이터를 상기 회귀 레이어에 입력하고, 상기 회귀 레이어의 출력을 바탕으로 변환 행렬을 추정할 수 있다.

[0012] 상기 회귀 레이어는, 상기 추정된 변환 행렬과 정답에 해당하는 변환 행렬 간의 MSE(Mean Square Error)를 바탕으로 구성된 손실 함수에 따라 훈련될 수 있다.

[0013] 한편, 상기 FPE(Frequency Positional Encoding) 레이어는, 상기 타겟 이미지에 대하여 1차 이상의 미분이 수행된 잔차(u'_h)에 대하여 2D FFT(Fast Fourier Transform)를 수행하여 $U_h(u, v)$를 획득하고, 상기 $U_h(u, v)$의 크기 성분을 일정 크기의 패치($x \in R^{\times 2}$)로 나누어 이하 수학식에 따른 시퀀스 z_0를 출력하고, $\mathbf{E} \in \mathbb{R}^{P^2 \times d}, \mathbf{E}_{pos} \in \mathbb{R}^{n \times d}$, $\mathbf{Z}_0 = [\mathbf{x}_p^1\mathbf{E}; \mathbf{x}_p^2\mathbf{E}; \ldots; \mathbf{x}_p^n\mathbf{E}] + \mathbf{E}_{pos}$, 상기 n은 패치의 결과 수이고, 상기 p는 패치의 크기이고, 상기 E_{pos}는 위치 인코딩에 해당할 수 있다.

[0014] 이때, 상기 트랜스포머 인코더는, 상기 시퀀스 z_0에 대하여 어텐션 연산을 수행하여 Q(Query), K(Key), 및 V(Value)를 획득하고, 인코딩된 입력 스펙트럼이 있는 Q와 스펙트럼 도메인에 나타날 것으로 예상되는 피크 정보가 있는 K 간의 내적을 통해 확률 값을 산출하고, 상기 확률 값을 V에 투영하여 추정과 관련된 데이터를 획득할 수 있다.

[0015] 상기 트랜스포머 인코더는, MSA(Multi-head Self-Attention) 레이어 및 MLP(Multi-Layer Perceptron) 레이어를 포함할 수 있다.

[0016] 본 개시의 일 실시 예에 따라 이미지에 적용된 변환을 추정하도록 훈련된 인공지능 모델을 포함하는 전자 장치의 제어 방법은, 상기 전자 장치가, 타겟 이미지를 상기 인공지능 모델에 입력하는 단계, 상기 전자 장치가, 상기 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정하는 단계를 포함한다. 상기 인공지능 모델은, 상기 타겟 이미지를 주파수 영역으로 변환하여 상기 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어, 상기 FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder)를 포함한다.

발명의 효과

[0017] 본 개시에 따른 전자 장치의 제어 방법은, 기존의 분류 기반 공식에서 탈피하여 변환 행렬을 직접 추정할 수 있는 인공지능 모델을 활용하였으며, 기존의 방법보다 오차가 적은 뛰어난 성능을 보였다.

도면의 간단한 설명

[0018] 도 1은 리샘플링과 관련된 다양한 작업을 거친 스펙트럼을 도시한 도면들,

도 2a는 본 개시의 일 실시 예에 따른 전자 장치의 구성을 설명하기 위한 블록도,

도 2b는 본 개시의 일 실시 예에 따라 변환 행렬을 추정하는 인공지능 모델의 아키텍처를 설명하기 위한 도면,

도 3은 본 개시의 일 실시 예에 따른 실험에 사용된 세 가지 테스트 세트의 샘플 패치들을 도시한 도면,

도 4는 본 개시에 따라 제안된 인공지능 모델에 의해 추정된 매개변수의 정오분류표를 도시한 도면,

도 5는 본 개시의 일 실시 예에 따른 각 매개변수의 RMSE 기반 비교 결과를 도시한 그래프들,

도 6은 본 개시의 일 실시 예에 따른 실험에서 압축에 대한 견고성 테스트 결과를 도시한 그래프들,

도 7은 본 개시의 일 실시 예에 따른 인공지능 모델의 전처리 레이어 및 트랜스포머 인코더의 마지막 레이어 각각의 활성화 맵을 도시한 도면들이다.

발명을 실시하기 위한 구체적인 내용

[0019] 본 개시에 대하여 구체적으로 설명하기에 앞서, 본 명세서 및 도면의 기재 방법에 대하여 설명한다.

[0020] 먼저, 본 명세서 및 청구범위에서 사용되는 용어는 본 개시의 다양한 실시 예들에서의 기능을 고려하여 일반적인 용어들을 선택하였다. 하지만, 이러한 용어들은 당해 기술 분야에 종사하는 기술자의 의도나 법률적 또는 기술적 해석 및 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 일부 용어는 출원인이 임의로 선정한 용어도 있다. 이러한 용어에 대해서는 본 명세서에서 정의된 의미로 해석될 수 있으며, 구체적인 용어 정의가 없으면 본 명세서의 전반적인 내용 및 당해 기술 분야의 통상적인 기술 상식을 토대로 해석될 수도 있다.

[0021] 또한, 본 명세서에 첨부된 각 도면에 기재된 동일한 참조번호 또는 부호는 실질적으로 동일한 기능을 수행하는 부품 또는 구성요소를 나타낸다. 설명 및 이해의 편의를 위해서 서로 다른 실시 예들에서도 동일한 참조번호 또는 부호를 사용하여 설명한다. 즉, 복수의 도면에서 동일한 참조 번호를 가지는 구성요소가 모두 도시되어 있다고 하더라도, 복수의 도면들이 하나의 실시 예를 의미하는 것은 아니다.

[0022] 또한, 본 명세서 및 청구범위에서는 구성요소들 간의 구별을 위하여 "제1", "제2" 등과 같이 서수를 포함하는 용어가 사용될 수 있다. 이러한 서수는 동일 또는 유사한 구성요소들을 서로 구별하기 위하여 사용하는 것이며 이러한 서수 사용으로 인하여 용어의 의미가 한정 해석되어서는 안 된다. 일 예로, 이러한 서수와 결합된 구성요소는 그 숫자에 의해 사용 순서나 배치 순서 등이 제한되어서는 안 된다. 필요에 따라서는, 각 서수들은 서로 교체되어 사용될 수도 있다.

[0023] 본 명세서에서 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "구성되다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0024] 본 개시의 실시 예에서 "모듈", "유닛", "부(Part)" 등과 같은 용어는 적어도 하나의 기능이나 동작을 수행하는 구성요소를 지칭하기 위한 용어이며, 이러한 구성요소는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, 복수의 "모듈", "유닛", "부(Part)" 등은 각각이 개별적인 특정한 하드웨어로 구현될 필요가 있는 경우를 제외하고는, 적어도 하나의 모듈이나 칩으로 일체화되어 적어도 하나의 프로세서로 구현될 수 있다.

[0025] 또한, 본 개시의 실시 예에서, 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 직접적인 연결뿐 아니라, 다른 매체를 통한 간접적인 연결의 경우도 포함한다. 또한, 어떤 부분이 어떤 구성요소를 포함한다는 의미는, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0026] 주기성 분석 방법을 설명하기 전에 이미지 신호와 기하학적 변환을 수학적으로 정의한다. 단계 크기 $\Delta \in R+$로 샘플링된 디지털 영상 신호 $u_0[n]$를 연속 신호 $u_h[x]$ 로 변환하기 위해 보간 필터 $h(x)$가 다음과 같이 사용될 수 있다.

수학식 1

[0027] $$u_h(x) = \sum_{i \in \mathbb{Z}^2} u_0[i] h\left(\frac{x}{\Delta} - i\right)$$

[0028] 매핑 함수 $h(x)$는 행렬 곱셈 형식으로 나타낼 수 있다. $y = Ax$. 회전 A_R 후에 적용할 스케일링 A_S를 갖는 변환 행렬 A는 다음과 같이 정의된다.

[0029] $$\mathbf{A} = \mathbf{A}_S \cdot \mathbf{A}_R = \begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} S\cos\theta & S\sin\theta \\ -S\sin\theta & S\cos\theta \end{bmatrix}$$

[0030] 여기서 S와 Θ는 각각 배율과 회전 각도를 나타낸다. 회전은 원본 이미지의 중심점을 기준으로 반시계 방향이라고 가정한다.

[0031] 보간(interpolation)의 주기성을 나타내기 위해 Andrew. C. Gallanher.는 보간된 신호의 2차 도함수를 적용하였고, 2차 도함수의 분산이 주기 Δ로 주기적임을 보여주었다. Mahdianet et al.은 차 도함수에 대한 일반화된 방법을 보여주었다.

수학식 2

[0032] $$D^n\{u_h\}(x) = \sum_{i=-\infty}^{\infty} u_0(i) D^n\{h\}\left(\frac{x}{\Delta} - i\right)$$

[0033] 여기서 D_n은 n차 도함수를 나타낸다. 위치 x의 함수로서 n번째 도함수의 분산은 다음과 같이 유도될 수 있다.

수학식 3

[0034] $$\mathrm{Var}\{D^n\{u_h\}(x + r\Delta)\} = \sigma^2 \sum_{i=-\infty}^{\infty} D^n\{h\}\left(\frac{x}{\Delta} - (i - r)\right)^2$$

수학식 4

[0035] $$\mathrm{Var}\{D^n\{u_h\}(x)\} = \mathrm{Var}\{D^n\{u_h\}(x + r\Delta)\}$$

[0036] 여기서 Var{·} 및 r은 각각 분산 및 정수이다. 따라서 보간된 신호와 그 파생물이 고유한 주기성을 갖는다는 결론이 도출될 수 있다(Weimin Wei et al.).

[0037] 보간의 주기는 스펙트럼 영역에서 피크 신호 분석을 통해 효과적으로 판별되었다(Babak Mahdian et al., Alin C Popescu et al.). 또한, 이미지의 기하학적 변형 정도 A로 표시되는 주기성 정보 이론이 연구되었다(Chenglong Chen et al.).

[0038] $[0, 1)^2$ 범위의 기저대역 스펙트럼을 기반으로, Chen et al.은 정규화된 주파수에서 나타나는 리샘플링 피크를 다음과 같이 표시하였다.

수학식 5

[0039] $$f_{\mathbf{A}}^{(m)} = \mathrm{frac}\left\{(\mathbf{A}^T)^{-1} m\right\}$$

[0040] 여기서 frac{·}은 요소별 소수 부분 톱니 함수를 나타내고 $m \in Z^2$이다.

[0041] 도 1은 리샘플링과 관련된 다양한 작업을 거친 스펙트럼을 보여준다. 각 이미지는 1000개의 고역 통과 필터링된 이미지에서 FFT 스펙트럼을 평균화하여 획득된 것이다. 이상적으로는 피크의 위치 정보에서 기하학적 변환을 유도할 수 있지만 실제 영상의 강한 내용 신호로 인해 기존 스펙트럼 기반 방법의 예측이 정확하지 않다.

[0042] 최근에는 다양한 영상 처리 분야에서 최첨단 성능을 보여준 CNN들도 이 문제에 적용되고 있다. 다층 컨볼루션 필터는 다양한 공간 주파수 대역에서 대역 통과 필터 역할을 하고 복잡한 기능을 학습하여 리샘플링 아티팩트를 추정한다. 그러나 기하 변환은 본질적으로 연속 공간에서 정의되기 때문에 분류 기반 방법은 고유한 조합의 수를 무한대로 확장할 수 없다는 고유한 한계가 있다. 따라서 이러한 연구는 스케일링 추정에 국한되었으며 예측에 대한 일반화된 솔루션을 제공하지 못했다.

[0043] 따라서, 본 개시는, 효율적인 파라미터 추정을 위해 FFT(Fast Fourier Transform)와 위치 부호화를 결합한 FPE(Frequency Positional Encoding) 레이어와, 주파수 영역에서 피크 위치 정보를 활용할 수 있는 엔드 투 엔드(end-to-end) 트랜스포머 기반의 추정기를 제안한다.

[0044] 본 개시는 이미지의 기하학적 변환 매개변수를 예측할 수 있는 엔드 투 엔드 트랜스포머 기반 추정기(estimator)를 제안한다. 고유한 리샘플링 기능을 사용하기 위해 고속 푸리에 변환 및 멀티 헤드 셀프 어텐션의 위치 인코딩을 사용하여 공간 주파수의 직접 분석이 수행된다. 회귀 레이어를 이전의 트랜스포머 인코더와 결합하면 이미지의 기하학적 변환 매개변수를 효과적으로 분석할 수 있다. 공개 데이터베이스를 통해 광범위한 비교 테스트를 수행한 결과, 본 개시에 따른 제어 방법은 대부분의 기존 방법보다 높은 예측 성능을 보여주고 압축(ex. JPEG)에 대한 높은 복원력도 보여준다.

[0045] 주파수 피크의 위치와 관계는 도 1과 같이 변환(transformation) 행렬 A를 계산하는 직접적인 단서 역할을 한다. 그러나, 많은 경우에 강한 이미지 콘텐츠 신호로 인해 리샘플링 주기의 독특한 피크를 분석하기 어렵다.

[0046] 따라서, 본 개시는, 주파수 영역의 주기성을 포함하여 다양한 신호 간의 관계를 캡처하여 이미지에 적용된 변환 행렬의 값을 추정하는 모델, 및 해당 모델을 이용하는 전자 장치를 제공한다. 이 경우 유도적 공간 편향이 있는 CNN과 달리 절대 공간 위치와 관련된 다양한 정보를 학습할 수 있기 때문에 트랜스포머 기반 모델이 더 적합하다.

[0047] 도 2a는 본 개시의 일 실시 예에 따른 전자 장치의 구성을 설명하기 위한 블록도이다. 도 2a를 참조하면, 전자 장치(100)는 메모리(110) 및 프로세서(120)를 포함한다.

[0048] 전자 장치(100)는 이미지의 위조 여부를 판단하거나 이미지를 복원하기 위한 적어도 하나의 장치 내지는 시스템으로 구현될 수 있다. 구체적으로, 전자 장치(100)는 서버로 구현될 수 있으며, 이밖에 이미지의 변환 여부/변환 정도를 판별하기 위한 소프트웨어/애플리케이션을 실행할 수 있는 데스크탑 PC, 노트북 PC, 태블릿 PC, 스마트폰 등 다양한 단말 기기에 해당할 수 있다.

[0049] 메모리(110)는 전자 장치(100)의 구성요소들의 전반적인 동작을 제어하기 위한 운영체제(OS: Operating System) 및 전자 장치(100)의 구성요소와 관련된 적어도 하나의 인스트럭션 또는 데이터를 저장하기 위한 구성이다.

[0050] 메모리(110)는 ROM, 플래시 메모리 등의 비휘발성 메모리를 포함할 수 있으며, DRAM 등으로 구성된 휘발성 메모리를 포함할 수 있다. 또한, 메모리(110)는 하드 디스크, SSD(Solid state drive) 등을 포함할 수도 있다.

[0051] 도 2a를 참조하면, 메모리(110)는 대상이 되는 타겟 이미지의 변환 여부/정도를 나타내는 변환 행렬을 추정하기 위한 인공지능 모델(111)을 포함할 수 있다.

[0052] 일 실시 예로, 인공지능 모델(111)은 타겟 이미지를 주파수 영역으로 변환하여 변환의 주기성을 인코딩하기 위한 FPE(Frequency Positional Encoding) 레이어(계층), FPE 레이어에서 출력된 스펙트럼 신호를 바탕으로 피크(peak)에 대한 위치 정보를 획득하는 트랜스포머 인코더(transformer encoder)를 각각 포함할 수 있다. 인공지능 모델(111)의 구성은 도 2b를 통해 보다 상세하게 후술한다.

[0053] 프로세서(120)는 전자 장치(100)를 전반적으로 제어하기 위한 구성이다. 구체적으로, 프로세서(120)는 메모리(110)와 연결되는 한편 메모리(110)에 저장된 적어도 하나의 인스트럭션을 실행함으로써 본 개시의 다양한 실시 예들에 따른 동작을 수행할 수 있다.

[0054] 프로세서(120)는 CPU, AP, DSP(Digital Signal Processor) 등과 같은 범용 프로세서, GPU, VPU(Vision

Processing Unit) 등과 같은 그래픽 전용 프로세서 또는 NPU와 같은 인공지능 전용 프로세서 등을 포함할 수 있다. 인공지능 전용 프로세서는, 특정 인공지능 모델의 훈련 내지는 이용에 특화된 하드웨어 구조로 설계될 수 있다.

[0055] 프로세서(120)는 타겟 이미지를 인공지능 모델(111)에 입력하여, 타겟 이미지에 적용된 변환에 매칭되는 변환 행렬을 추정할 수 있다.

[0056] 한편, 상술한 메모리(110) 및 프로세서(120) 외에도, 전자 장치(100)는 사용자 입력부, 통신부 등을 더 포함할 수 있다.

[0057] 사용자 입력부는, 사용자로부터 다양한 명령 내지는 정보를 입력 받기 위한 구성이다. 전자 장치(100)는 사용자 입력부를 통해 수신된 사용자 입력에 따라 (복수의 이미지 중) 타겟 이미지를 선택하거나, 타겟 이미지에 대한 감식(: 변환 여부, 변환 정도에 대한 판단)을 수행할 수 있다.

[0058] 사용자 입력부는 키패드, 터치패드, 마이크(: 음성 인식), 카메라(: 이미지 내 텍스트 인식) 등을 포함할 수 있다. 또한, 전자 장치(100)는 적어도 하나의 사용자 입력 장치(ex. 마우스, 키보드)와 연결되어 사용자 입력 장치를 통해 수신되는 사용자 입력에 따라 다양한 정보 내지는 명령을 획득할 수도 있다.

[0059] 통신부는, 전자 장치(100)가 적어도 하나의 외부 장치와 통신을 수행하기 위한 구성으로 회로를 포함할 수 있다.

[0060] 통신부는 유선 통신 및/또는 무선 통신을 통해 구현된 네트워크를 기반으로, 외부 장치와 연결될 수 있다. 이때, 통신부는 외부 장치와 직접적으로 연결될 수도 있지만, 네트워크를 제공하는 하나 이상의 외부 서버(ex. ISP(Internet Service Provider))를 통해서 외부 전자 장치와 연결될 수도 있다.

[0061] 네트워크는 영역 또는 규모에 따라 개인 통신망(PAN; Personal Area Network), 근거리 통신망(LAN; Local Area Network), 광역 통신망(WAN; Wide Area Network) 등일 수 있으며, 네트워크의 개방성에 따라 인트라넷(Intranet), 엑스트라넷(Extranet), 또는 인터넷(Internet) 등일 수 있다.

[0062] 무선 통신은 LTE(long-term evolution), LTE-A(LTE Advance), 5G(5th Generation) 이동통신, CDMA(code division multiple access), WCDMA(wideband CDMA), UMTS(universal mobile telecommunications system), WiBro(Wireless Broadband), GSM(Global System for Mobile Communications), DMA(Time Division Multiple Access), WiFi(Wi-Fi), WiFi Direct, Bluetooth, NFC(near field communication), Zigbee 등의 통신 방식 중 적어도 하나를 포함할 수 있다.

[0063] 유선 통신은 이더넷(Ethernet), 광 네트워크(optical network), USB(Universal Serial Bus), 선더볼트(ThunderBolt) 등의 통신 방식 중 적어도 하나를 포함할 수 있다. 여기서, 통신부(150)는 상술한 유무선 통신 방식에 따른 네트워크 인터페이스(Network Interface) 또는 네트워크 칩을 포함할 수 있다. 한편, 통신 방식은 상술한 예에 한정되지 아니하고, 기술의 발전에 따라 새롭게 등장하는 통신 방식을 포함할 수 있다.

[0064] 일 실시 예로, 전자 장치(100)는 통신부를 통해 외부 전자 장치(ex. 외부 서버, 데이터 베이스, 외부 단말 등)와 연결되어, 타겟 이미지 또는 상술한 인공지능 모델(111)의 적어도 일부를 수신할 수 있다. 또한, 전자 장치(100)는 전자 장치(100)가 수행한 감식의 결과를 외부 전자 장치로 전송할 수도 있다.

[0065] 도 2b는 본 개시의 일 실시 예에 따라 변환 행렬을 추정하는 인공지능 모델의 아키텍쳐를 설명하기 위한 도면이다.

[0066] 도 2b를 참조하면, 본 개시의 일 실시 예에 따른 인공지능 모델(111)은 전처리 레이어(계층), 주파수 위치 인코딩 레이어(계층), 트랜스포머 인코더, 및 회귀 레이어(계층)의 네 가지 주요 모듈로 구성될 수 있다.

[0067] (1) 전처리 레이어(Preprocessing layer)

[0068] 이미지의 지배적인 콘텐츠 신호는 필터링되어 주기적 보간 궤적을 나타낸다. 기대 최대화(Alin C Popescu and Hany Farid et al.), 미분 필터(Babak Mahdian et al., Seung-Jin Ryu et al.), 훈련 가능한 컨볼루션 필터(Belhassen Bayar et al., Mehdi Boroumand et al.)를 기반으로 원본 이미지에서 리샘플링 트레이스를 감지하기 위해 다양한 기술이 도입되었다. 본 개시의 일 실시 예에 따른 인공지능 모델(111)은, 실증적으로 최고의 성능을 보인 Laplacian 필터와 함께 CNN 기반 필터링 수단을 포함하는 전처리 레이어를 포함할 수 있다.

[0069] 전처리 레이어를 통한 프로세서(120)의 동작을 설명하면, 먼저, 입력 계조 영상에 라플라시안 필터를 적용하여

2차 미분 결과를 얻는다. 필터링된 이미지에서 풍부한 공간 정보를 얻기 위해 16개 출력 채널을 가진 제한된 컨볼루션 레이어로 구성된 특징 확장 레이어가 적용된다. 마지막으로, 잔여 신호 집계를 위한 단일 출력 채널을 갖는 7 × 7 컨볼루션 레이어가 적용된다. 전처리 레이어는 이미지가 주파수 분석을 위한 최적의 형태로 필터링되도록 훈련되는 것을 목표로 한다.

[0070] (2) 주파수 위치 인코딩 레이어(Frequency positional encoding layer)

[0071] 본 개시에 따른 인공지능 모델(111)은, 트랜스포머의 분리된 입력 패치로 변환(transformation) 내지는 보간(interpolation)의 명시적 주기성을 인코딩하기 위해 FPE(Frequency Positional Encoding) 레이어를 포함할 수 있다. 컨볼루션 이론(the convolution theorem)에 따르면 공간 도메인의 컨볼루션은 주파수 도메인의 포인트 별 곱셈과 같다. 따라서 주파수 영역의 학습은 CNN(convolution filter-based learning method)과 동일한 결과를 얻는다.

[0072] 위치 인코딩에 따라, 공간 영역에서 나타나는 피크의 절대 위치가 식별될 수 있다.

[0073] 주파수 영역 학습에 대해 유사한 개념을 사용한 다른 연구(Yongming Rao et al., Kai Xu et al.)와 달리 기하학적 변환 문제의 추정은 주파수 영역에서 명시적 피크를 갖는 것으로 알려져 있다(Babak Mahdian et al.). 그러므로 트랜스포머를 사용한 주파수 영역의 명시적 학습이 매개변수 요구 사항 및 추정 성능 측면에서 도움이 될 수 있다.

[0074] 전처리된 잔차 ′ = {} ∈ R× 가 주어지면, 공간 차원에 따라 2D FFT가 수행되어 이미지를 주파수 영역으로 변환될 수 있다.

수학식 6

[0075] $$U_h(u,v) = \mathcal{F}[u'_h[x,y]] \in \mathbb{C}^{H\times W}$$

[0076] 여기서 F는 2D FFT를 나타낸다. FFT 출력은 켤레 대칭이며, 따라서 음의 주파수 항은 중복되어 최종 추정에 영향을 미칠 수 있다.

[0077] 그 후, 실제 값 입력에 FFT가 적용되어 $U_h \in C^{H\times W''}$ 를 출력할 수 있다. 여기서, $W' = \lfloor W/2 \rfloor + 1$

[0078] 위치 인코딩은 스펙트럼 피크 위치를 학습 가능하게 하며, 이는 본 개시에 따른 해결 수단에 있어 본질적인 것이다. 이미지의 전체 영역의 위치를 사용하기 위해 절대 위치 인코딩(Jean-Baptiste Cordonnier et al.)이 사용될 수 있다. 먼저 잠재적인 리샘플링 트레이스의 스펙트럼을 나타내기 위한 크기 성분 $\bar{U}_h = |U_h|$이 도출될 수 있다. 패치 인코딩은 크기 $\bar{U}_h$ 를 일련의 평평한 패치 $x_p \in R^{n\times p^2}$ 로 나눌 수 있다. 여기서 n은 패치의 결과 수이다(n = HW/P^2).

수학식 7

[0079] $$\mathbf{E} \in \mathbb{R}^{P^2\times d},\ \mathbf{E}_{pos} \in \mathbb{R}^{n\times d}$$

수학식 8

[0080] $$\mathbf{Z}_0 = [\mathbf{x}_p^1\mathbf{E}; \mathbf{x}_p^2\mathbf{E}; \ldots; \mathbf{x}_p^n\mathbf{E}] + \mathbf{E}_{pos}$$

[0081] 여기서 패치 크기 = 16이 경험적으로 선택될 수 있다. 출력 크기는 패치 크기 P의 정수 배수여야 하므로 선행 열이 제거될 수 있다. 패치 인코딩 크기와 동일한 위치 정보가 수학식 8의 패치 표현에 추가된다. 트랜스포머는 이미지 내 패치 사이의 상대적 거리를 학습하고 더 가까운 패치는 더 유사한 인코딩 매개변수를 갖는 경향이 있

다.

[0082] 인코딩된 입력 데이터는 이미지의 각 패치에 대한 위치 정보를 모르기 때문에 학습 가능한 위치 인코딩(E_{pos})은 각 시퀀스의 위치 정보를 제공한다. 위치 인코딩을 기반으로 이미지의 공간 정보를 학습할 수 있으며 모든 컨볼루션 레이어를 멀티 헤드(multi-head) 셀프 어텐션 레이어(self-attention layers)로 대체할 수 있다.

[0083] (3) 트랜스포머 인코더(Transformer encoder)

[0084] 필터링된 스펙트럼 신호에서 분명한 피크 위치 정보를 획득하기 위해 트랜스포머에 대한 self-attention 메커니즘(Alexey Dosovitskiy et al.)이 채택되었다.

[0085] 트랜스포머 인코더는 이미지 내의 패치 사이의 상대 거리를 학습하므로 어텐션 레이어가 피크 위치에 집중할 수 있다.

[0086] 본 개시에 따라 제안된 인공지능 모델은 Attention 연산을 위한 전체 입력 정보를 사용한다. 네트워크의 깊이가 증가함에 따라 어텐션 거리가 증가하고 CNN의 수신 필드와 유사한 글로벌 영역을 효과적으로 캡처한다. 입력 시퀀스 0은 먼저 3개의 학습 가능한 행렬에 투영되어 Q(Query), K(Key) 및 V(Value)를 얻는다. 셀프 어텐션 레이어의 출력 $Z \in R^{n \times d_v}$ 은 다음과 같다.

수학식 9

[0087] $$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\mathbf{Q}\mathbf{K}_T / \sqrt{d_q}\right)\mathbf{V}$$

[0088] 인코딩된 입력 스펙트럼이 있는 Q와 스펙트럼 도메인에 나타날 것으로 예상되는 피크 정보가 있는 K 간의 내적을 통해 각 입력은 입력 값과 연관성이 정의될 수 있다. Q의 차원 d_q로 scale을 조정한 후, softmax 함수를 적용하여 확률 값을 구할 수 있다. 그런 다음 확률 값을 V에 투영하여 이미지 기하 변환에 대한 추정 단서를 추출할 수 있다. 입력의 서로 다른 요소에서 여러 관계를 캡슐화하기 위해 멀티 헤드(multi-head) 셀프 어텐션(self-attention)은 복수의 셀프 어텐션(self-attention) 블록으로 구성될 수 있다.

수학식 10

[0089] $$\mathbf{Z}_0^{\text{MSA}} = \text{MSA}(\text{LN}(\mathbf{Z}_0)) + \mathbf{Z}_0$$

수학식 11

[0090] $$\mathbf{Z}_1 = \text{MLP}(\text{LN}(\mathbf{Z}_0^{\text{MSA}})) + \mathbf{Z}_0^{\text{MSA}}$$

수학식 12

[0091] $$\mathbf{Z}_\ell = \text{MLP}(\text{LN}(\mathbf{Z}_{\ell-1}^{\text{MSA}})) + \mathbf{Z}_{\ell-1}^{\text{MSA}}, \quad \ell = 1 \ldots L$$

[0092] 트랜스포머 인코더에서 MSA(multi-head self-attention) 레이어와 MLP(multi-layer-perceptron)는 교대로 구성된다. 헤드 수 = 12로 설정될 수 있다. 레이어 정규화(LN) 및 잔여 연결이 모든 레이어에 적용될 수 있다. 인코더 블록은 번 연속으로 반복된다(수학식 12).

[0093] (4) 회귀 레이어(Regression layer)

[0094] 기존의 분류 기반(Belhassen Bayar et al., Anotol Maier et al.) 접근 방식은 고유한 조합의 수를 무한정 확

장할 수 없다는 고유한 한계가 있다. 따라서, 타겟이 되는 변환 매개변수를 직접 추정하기 위해 회귀 레이어(regression layer)가 도입되었다. 회귀 기반 접근 방식은 일반화된 추정을 위한 더 나은 방법이다. 또한, 본 개시에 따른 모델의 레이어는 CNN 기반 접근 방식과 같은 기존의 기초 방식과 결합될 수 있다.

[0095] 수학식 12의 결과 Z_ℓ를 기반으로 GAP(Global Average Pooling)를 수행한 후 추정된 변환 행렬 $\hat{A} \in R^{2\times 2}$는 다음과 같이 획득될 수 있다.

수학식 13

[0096] $$\hat{\mathbf{A}} = \mathrm{MLP}(\mathrm{LN}(\mathrm{GAP}(\mathbf{Z}_\ell)))$$

[0097] 회귀 계층의 경우 GELU 활성화 함수가 있는 MLP가 사용되었다. 제안된 모델을 최적화하기 위해 $\hat{A}$와 정답(ground truth) 레이블 A 사이의 평균 제곱 오차(MSE)가 손실 함수로 사용될 수 있다. Min Lin et al.에서 논의된 GAP를 고려했을 때, MLP 계층의 구성은 중간에 병목 계층을 배치하지 않는 한 추정 성능에 큰 영향을 미치지 않습니다. 마지막으로 회전도 Θ와 스케일링 인자 S 같은 변환 인자는 $\hat{A}$의 분해로부터 도출될 수 있다.

[0098] 이하 서술을 통해서는, 본 개시에 따른 상술한 인공지능 모델(111)의 성능을 확인하기 위한 실험의 내용 및 결과에 대해 설명한다.

[0099] 엄격한 실험을 위해서는 이미지 데이터 세트의 처리 이력을 보장하는 것이 중요하다. 공개적으로 사용 가능한 원시 이미지 데이터 세트 RAISE-6k[2]를 사용했다. 훈련 데이터 세트와 테스트 데이터 세트는 4:1 비율로 분할되었다. 각 이미지는 기본 이미지 획득에서 CFA(컬러 필터 어레이) 보간으로 인한 간섭 신호를 피하기 위해 2의 인수로 다운샘플링된다. 데이터 세트의 각 이미지에서 10개의 이미지 패치(512 × 512 해상도)를 무작위로 잘라냈습니다. 결과적으로 47,985개의 이미지 패치를 얻었다.

[0100] 스케일링 계수 s = 0.5 + k*0.05가 사용되었다. 여기서 k는 10을 제외하고 0과 29 사이에서 무작위로 선택된다. 회전은 각도 Θ = {1°, ..., 45°}가 사용되었다. 각도에 대한 주파수 쌍이 스펙트럼 영역에서 대칭이기 때문에 각도 범위 Θ는 45°로 제한된다. 도 3과 같이 rotation only(R), scaling only(S), mixed(rotation+scaling) transformation-based (M) evaluation의 세 가지 테스트 데이터 세트가 수집되었다. 테스트 데이터 세트의 기하학적 변환 매개변수는 훈련 데이터 세트와 동일한 범위 및 간격으로 구성되었다. 훈련하는 동안 기하 변환은 모든 에포크에서 동시에 적용되었다. 동일한 이미지를 생성하는 단위 행렬에 기반한 변환은 손실 수렴을 방해하는 것을 방지하기 위해 실험에서 제거되었다. 이미지의 직사각형 테두리는 예측을 위해 강력하고 불필요한 신호를 수반하므로 변환된 패치의 중심이 256 × 256 해상도로 다시 잘린다.

[0101] 본 실험에 따라, 본 개시의 일 실시 예에 따른 인공지능 모델의 알고리즘은 Pytorch 라이브러리를 사용하여 구현되었다. Intel Xeon(R) Gold 6226R(2.90GHz) GPU 기반 머신과 2개의 RTX 3090 그래픽 카드가 교육 및 평가에 사용되었다. 수행된 모든 실험에 대해 배치 크기는 32로 설정되었으며 학습률 l = 4 · 10 − 5인 cosine annealing warm restart scheduler와 AdamW 최적화 프로그램이 사용되었다. 제안된 네트워크가 수렴하는 데 약 200에포크(8시간)의 학습이 필요했다. 데이터 증강을 이미지에 적용할 때 리샘플링 아티팩트가 발생할 수 있으므로 정규화를 제외한 데이터 증강을 적용하지 않았다.

[0102] 이 논문에서 예측된 결과는 두 가지 형식이 있다: 추정된 매개변수 $\hat{S}$, $\hat{\theta}$ 및 행렬 $\hat{A}$. 각 평가에 대해 다음 측정 항목을 사용하는 것이 좋다. 먼저 추정된 매개변수 $\hat{S}$ 및 $\hat{\theta}$ 의 정확도를 평가하기 위해 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE) 및 결정 계수(R^2 점수)의 세 가지 측정항목이 사용되었다. 결정 계수는 다음과 같이 정의된다.

수학식 14

[0103] $$R^2(\mathbf{y},\hat{\mathbf{y}}) = 1 - \frac{\sum_t (y_t - \hat{y}_t)}{\sum_t (y_t - \bar{y})}$$

[0104] 여기서 $\bar{y}$ 는 레이블의 평균을 나타낸다. 결합 변환 S+R을 평가할 때 R 및 S의 범위가 다르므로 편향된 점수가 발생할 수 있다. 불균형 요인을 완화하기 위해 거리 점수를 측정하기 전에 R 요인을 라디안으로 변환하였다.

[0105] 추정된 $\hat{\mathbf{A}}$ 사이의 차이의 크기를 측정하려면, 변환 행렬 **A**에서 F-norm으로 표시된 Rami Hagege et al.의 거리 메트릭을 채택할 수 있다. 이는 다음과 같이 정의될 수 있다.

수학식 15

[0106] $$F(\mathbf{A},\hat{\mathbf{A}}) = \frac{1}{2}||\mathbf{A}^{-1}\hat{\mathbf{A}} - \mathbf{I}||_F$$

[0107] $||\cdot||$ 및 **I**는 각각 Frobenius norm 및 단위 행렬을 나타낸다.

표 1

[0108] **Table 1: Evaluation ($\mathcal{M}$) result of the estimated matrix $\hat{\mathbf{A}}$ using the preprocessing layer.**

[0109]

Preprocessing layer				
Laplacian	Constrained	Conv	*RMSE*	*F-norm*
-	-	-	.0624	.0405
-	-	✓	.0720	.0488
-	✓	✓	.0668	.0454
✓	-	-	.0610	.0392
✓	✓	✓	.0583	.0372

Constrained: 16@5x5 [1], Conv: 1@7x7

표 2

[0110] **Table 2: Comparison of transformer encoder configuration on the test set $\mathcal{M}$.**

[0111]

Transformer encoder						
P	d_{pos}	N_h	D	# param.	*RMSE*	*F-norm*
8	1D	12	12	10.55M	.0506	.0316
16	1D	8	12	28.31M	.0583	.0373
16	1D	12	8	30.34M	.0547	.0355
16	1D	12	12	42.34M	.0481	.0300
16	2D	12	12	42.34M	.0625	.0406
32	1D	12	12	172.34M	.1097	.0851

P: patch size, d_{pos}: positional encoding dim, N_h: attention num head, D: encoder depth

[0112] 많은 연구에서 이미 전처리 레이어의 효과가 입증되었다. 표 1과 같이 전처리 레이어에 가장 적합한 조합을 경험적으로 선택하기 위해 ablation 실험이 수행되었다. 나머지 모듈은 기본 구조를 따른다. 여기에서 5x5 Laplacian 필터, 16개 출력 채널을 가진 제약 필터 및 단일 출력 채널을 가진 7x7 컨볼루션 레이어가 사용되었다. 이미지의 지배적인 내용 정보 때문에 훈련은 전처리 없이 수렴되지 않았다. 흥미롭게도 Laplacian과 Constrained Filter를 결합하면 역할이 다소 겹치더라도 일관되게 추정이 향상되었다. 결과적으로, 실험의 나머

지 부분에서 가장 성능이 좋은 전처리 모듈 구조가 선택되었다.

[0113] 베이스라인 트랜스포머 인코더의 구조에 대한 실험이 수행되었다. 실험은 패치 크기(P), 위치 인코딩 차원(d_{pos}), 헤드 수(N_h) 및 인코더 깊이(D) 등 인코더의 매개변수의 구성을 변경하여 진행되었다. 추출된 특징 맵의 크기는 패치 크기가 증가함에 따라 감소하며, 이는 MSA 및 MLP 매개변수의 양에 영향을 미친다. 위치 인코딩 행렬 E_{pos}에 대해, 1D 및 2D로 표시된 두 가지 방법을 테스트했으며, 1) 패치 인코딩을 거친 후 시퀀스로 위치 정보를 인코딩(1D)하고 2) 위치 정보를 2차원 형식으로 임베딩(2D)하였다. 표 2와 같이 가장 성능이 좋은 P = 16, N_h = 12, D = 12를 실증적으로 선택하였다. 1차원 위치 부호화가 2차원 부호화보다 더 좋은 점수를 보였다.

[0114] {256, 4}, {256, 200, 4}, {256, 200, 100, 4}, {256, 200, 200, 100, 4}와 같이 MLP 레이어의 수를 늘리면서 다양한 실험을 수행했다. 여기서 {in, out} 은 GELU 활성화가 있는 in · out 크기의 선형 레이어를 나타낸다. RMSE 및 H 메트릭 기반 평가에서 단일-레이어 및 3-레이어 MLP는 각각 0.0583 및 0.0370에서 가장 높은 성능을 보였다. 이전 연구에서 논의된 바와 같이 선형 레이어의 수가 추정 성능에 큰 영향을 미치지 않는다는 것을 관찰했다. 따라서 모든 기준 실험에서 단일-레이어 {256, 4} 퍼셉트론을 사용했다. 도 4는 제안된 네트워크의 추정된 매개변수의 정오분류표를 보여준다. 매개변수는 추정된 행렬 $\hat{A}$의 분해에서 파생되었다. 테스트 세트 S에서는 1.0 스케일링을 제외하고 전체 스케일링 계수에 따라 성능 차이가 관찰되지 않았다. 일반적으로 보간 퀘적이 약한 다운 샘플링 변환에 대해 원본 이미지를 식별할 수 있기 때문이다. S = 1.0 및 분해된 S(from M)에 대한 추정은 스케일링 인자 S보다 더 정확하다.

[0115] S에 대한 추정과 달리 스케일링 인자 S는 회전과 결합되어 변환 결과가 더 일관된 리샘플링 추적이 되도록 한다.

[0116] 기존 연구들 중 총 2개의 수작업 기반 방법(Chenglong Chen et al., Seung-Jin Ryu et al.)과 3개의 CNN 기반 분류 모델(Belhassen Bayar et al., Chang Liu et al., Lin Peng.)을 사용하여 비교 실험을 수행했다. Chen et al.은 임계값 기반 피크 검출기를 사용하여 중요한 피크를 필터링하고 선택한 피크의 위치에서 회전 계수를 추정했다. Chenglong Chen et al. 은 회전 계수만을 추정하였기 때문에 평가에는 R과 R+S만을 사용하였다. CNN 기반 방법은 스케일링된 데이터로만 학습한 기존 연구와 달리 제안된 모델과 동일한 데이터를 사용하여 학습되었다. Chang Liu et al.은 64 × 64의 고정 크기 입력을 받도록 설계되었으므로 입력 크기가 변경되면 네트워크의 전체 구조가 변경되어야 한다. 따라서 원본 논문과 동일한 설정을 사용하여 Chang Liu et al.에 대한 실험을 수행했다. CNN 기반 연구의 인코더는 논문에 제시된 것과 동일한 매개 변수를 사용했지만 마지막 MLP 계층의 출력 수는 평가 데이터 세트(S: 30, R: 46 및 M: 1379(=30×46-1))에 따라 변경되었다.

표 3

[0117] **Table 3: Overall evaluation results of the proposed and compared methods.**

	Test set $\mathcal{R}$			Test set $\mathcal{S}$			Test set $\mathcal{M}$				Test set $\mathcal{M}'$			
	RMSE	*MAE*	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2	*F-norm*	*RMSE*	*MAE*	R^2	*F-norm*
Ryu et al.[24]	.3327	.2882	-1.5849	.6075	.5683	-1.4996	.3259	.2826	-1.4203	.4398	.5474	.5474	-1.4788	.5915
Chen et al.[4]	.2045	.0964	-1.2084	-	-	-	.5865	.4922	-6.4052	.3482	.6100	.5121	-6.8109	.3483
Bayar et al.[2]	.1811	.1139	.2650	.1560	.0990	.8450	.1496	.0795	.9108	.0889	.1505	.0847	.9138	.0955
Liu et al.[18]	.0485	.0132	.9066	.0930	.0308	.9295	.0721	.0208	.9688	.0269	.1819	.1015	.8816	.1064
Peng et al.[21]	.0389	.0134	.9272	.0741	**.0297**	.9554	.0508	.0168	.9836	.0211	**.0604**	**.0273**	.9810	**.0305**
Ours	**.0174**	**.0087**	**.9846**	**.0652**	.0300	**.9684**	**.0301**	**.0142**	**.9946**	**.0163**	.0620	.0309	**.9844**	.0355
Ablation experiments follow:														
Prefilter	FPE	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*	R^2	*F-norm*	*RMSE*	*MAE*	R^2	*F-norm*
-	-	.1221	.0890	.6811	.1136	.0812	.9199	.1106	.0771	.9542	.0912	.1064	.0745	.9589	.0858
✓	-	.0609	.0378	.9072	.0757	.0481	.9627	.0600	.0337	.9852	.0377	.0599	.0343	.9859	.0381
-	✓	.0764	.0487	.8629	.0850	.0558	.9535	.0682	.0390	.9813	.0455	.0644	.0378	.9839	.0433
✓	✓	.0174	.0087	.9846	.0652	.0300	.9684	.0301	.0142	.9946	.0163	.0620	.0309	.9844	.0355

$\mathcal{R}$: only rotation, $\mathcal{S}$: only scaling, $\mathcal{M}$: mixed(rotation+scaling), $\mathcal{M}'$: continuous-valued mixed(rotation+scaling)

[0118]

[0119] 표 3은 세 가지 테스트 세트의 제안 및 비교 방법에 대한 결과를 종합적으로 평가한 것이다. Chen et al.은 R 테스트 세트에서만 좋은 성능을 나타내도록 회전 계수를 추정하는 방법을 제시했다. MISLnet의 평가 결과는 딥러닝 기반 실험 중 상대적으로 낮은 성능을 보였다. 특히 M 테스트 세트를 적용했을 때 성능이 크게 저하되었다. 전체 매개변수의 수가 다른 네트워크에 비해 상당히 적기 때문에 MISLnet은 기하 변환의 일반적인 추정을 위한 인코더 용량이 충분하지 않았다.

[0120] - Ablation study

[0121] ablation 연구에서 제안된 전처리 계층과 FPE 계층의 효율성을 관찰하기 위해 본 개시에 따라 제안된 네트워크를 재구성했다. 표 3의 세 번째 섹션에서 볼 수 있듯이 전처리 계층은 모든 경우에 전반적인 성능을 향상시킨다. FPE 모듈은 프리필터만 사용한 경우에도 유사한 개선을 보였다. FPE 레이어는 전처리 레이어의 어떤 조합으로도 달성할 수 없는 회전 테스트 세트 R의 성능을 향상시켰다. 두 계층을 모두 적용했을 때 M의 RMSE는 0.0481로 제안된 프레임워크의 가장 좋은 성능을 보였다.

[0122] - Continuous valued test set M′

[0123] 임의의 변환 매개변수를 통해 조작하는 것은 실제 세계에서 보다 일반적인 시나리오이며 문제를 복잡하게 만든다. 시나리오의 효과를 확인하기 위해 추가 테스트 세트 M'을 사용하여 평가를 수행했다(표 3의 5번째 열 참조). 테스트 세트 M'의 이미지 패치는 훈련 데이터 세트에 사용된 연속 범위에서 소수점 이하 4자리 범위 내에서 무작위 샘플링 값을 변환하여 생성되었다. 제한된 수의 클래스를 가져야 하는 분류 공식을 기반으로 하는 매개변수 추정은 근본적인 한계가 있다. 중간 기능은 상당히 정확하지만 결과는 교육 과정에서 지정된 레이블에서 가장 가까운 클래스에 매핑되어야 한다. 그러나 본 개시에 따라 제안된 방법은 테스트 세트 M'에서 성능 저하를 초래하는 훈련 세트에도 과적합되었다. 따라서 회귀 계층과 훈련 전략에 대한 추가적인 연구가 필요하다.

[0124] - Robustness to JPEG compression

[0125] 실제 시나리오에서 이미지를 저장하고 전송할 때 JPEG 압축은 고주파 성분에 심각한 손상을 일으키는 일반적으로 사용되는 알고리즘이다. 품질 계수 $\in \{99, 98, 95, 90, 85, 80\}$를 사용하여 JPEG 압축에 대한 견고성 테스트를 수행했다. 각 이미지는 지정된 기하학적 변환을 적용한 후 압축되었다. 본 개시에 따라 제안된 네트워크(인공지능 모델)는 사전 훈련된 베이스라인 모델을 기반으로 미세 조정되었다. 도 6은 S, R, M에서 추정된 각 요인에 대한 유도된 변화를 나타낸다. 요약하면 본 개시에 따라 제안된 네트워크는 JPEG의 약한 수준에서 약간의 견고성을 보여주었다.

[0126] - Activation pattern analysis

[0127] 제안된 프레임워크가 리샘플링 아티팩트를 처리하는 방법을 이해하기 위해 활성화 맵 시각화를 기반으로 한 내부 표현을 관찰하였다(Ramprasaath R Selvaraju et al.). 중간 트랜스포머 네트워크에서 사용되지 않는 음의 주파수 용어도 시각화되었다. 도 7은 본 개시에 따른 제어 방법의 각 계층의 활성화 맵을 보여준다. 흥미롭게도 전처리 레이어는 스트라이프 패턴으로 표시되는 리샘플링 주기가 잘 보존된 이미지의 활성화를 보여준다. 다음 FFT 레이어에서 활성화 맵은 주파수 영역에서 관찰되었으며 주요 피크에 초점을 맞추었다(: 타겟 기하학 변환을 나타낸다). 분명히, 본 개시에 따라 제안된 트랜스포머 모듈은 주파수 피크의 위치 정보를 기반으로 매개변수를 추정하도록 최적화되었다.

[0129] 한편, 이상에서 설명된 다양한 실시 예들은 서로 저촉되지 않는 한 복수의 실시 예가 결합되어 구현될 수 있다.

[0130] 한편, 이상에서 설명된 다양한 실시 예들은 소프트웨어(Software), 하드웨어(Hardware) 또는 이들의 조합된 것을 이용하여 컴퓨터(Computer) 또는 이와 유사한 장치로 읽을 수 있는 기록 매체 내에서 구현될 수 있다.

[0131] 하드웨어적인 구현에 의하면, 본 개시에서 설명되는 실시 예들은 ASICs(Application Specific Integrated Circuits), DSPs(Digital Signal Processors), DSPDs(Digital Signal Processing Devices), PLDs(Programmable Logic Devices), FPGAs(Field Programmable Gate Arrays), 프로세서(Processors), 제어기(Controllers), 마이크로 컨트롤러(Micro-controllers), 마이크로 프로세서(Microprocessors), 기타 기능 수행을 위한 전기적인 유닛(Unit) 중 적어도 하나를 이용하여 구현될 수 있다.

[0132] 일부의 경우에 본 명세서에서 설명되는 실시 예들이 프로세서 자체로 구현될 수 있다. 소프트웨어적인 구현에 의하면, 본 명세서에서 설명되는 절차 및 기능과 같은 실시 예들은 별도의 소프트웨어 모듈들로 구현될 수 있다. 상술한 소프트웨어 모듈들 각각은 본 명세서에서 설명되는 하나 이상의 기능 및 작동을 수행할 수 있다.

[0133] 한편, 상술한 본 개시의 다양한 실시 예들에 따른 전자 장치에서의 처리동작을 수행하기 위한 컴퓨터 명령어(Computer Instructions) 또는 컴퓨터 프로그램은 비일시적 컴퓨터 판독 가능 매체(Non-transitory Computer-readable Medium)에 저장될 수 있다. 이러한 비일시적 컴퓨터 판독 가능 매체에 저장된 컴퓨터 명령어 또는 컴퓨터 프로그램은 전자 장치 등 특정 기기의 프로세서에 의해 실행되었을 때 상술한 다양한 실시 예에 따른 전자 장치의 동작을 상술한 특정 기기가 수행하도록 한다.

[0134] 비일시적 컴퓨터 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(Reading)이 가능한 매체를 의미한다. 비일시적 컴퓨터 판독 가능 매체의 구체적인 예로는, CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM 등이 있을 수 있다.

[0135] 이상에서는 본 개시의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 개시는 상술한 특정의 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 개시에 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안 될 것이다.

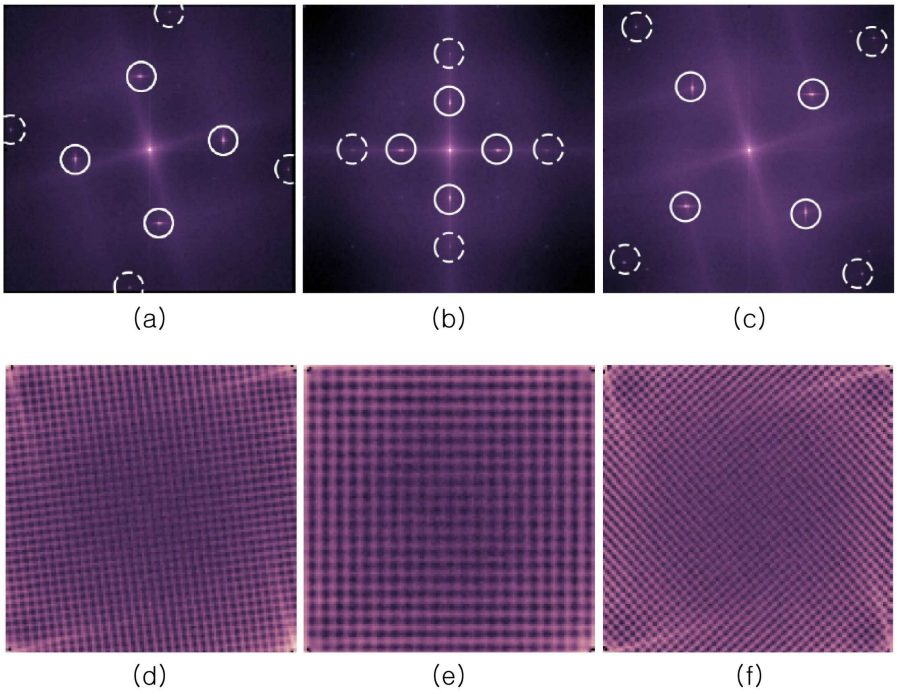
부호의 설명

[0136] 100: 전자 장치 110: 메모리

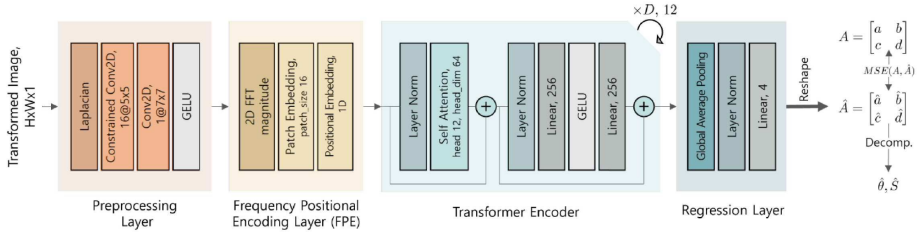
120: 프로세서

도면

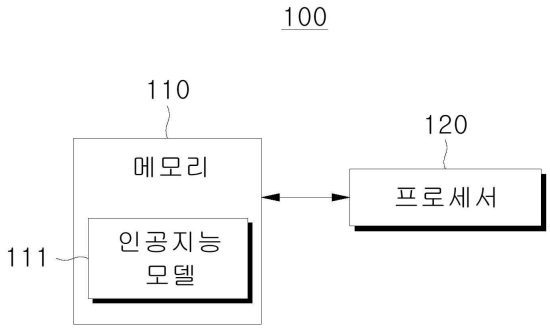
도면1



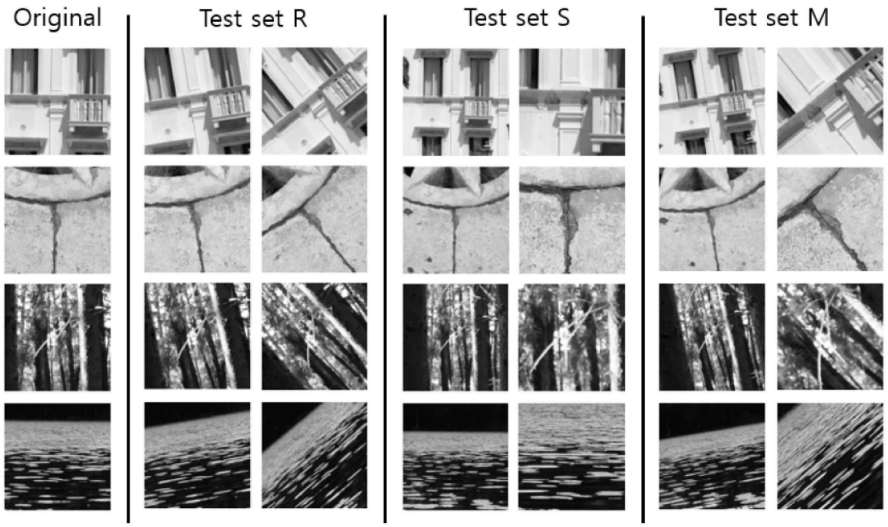
도면2a



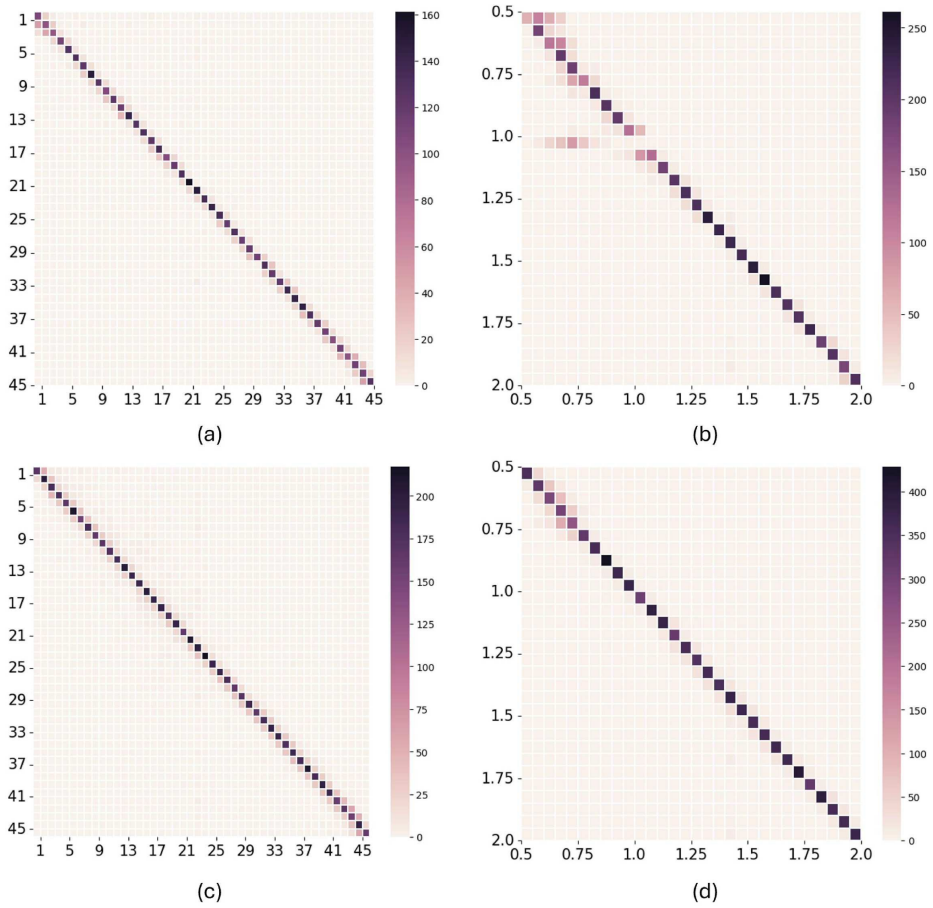
도면2b


100
110
메모리
120
프로세서
111
인공지능
모델

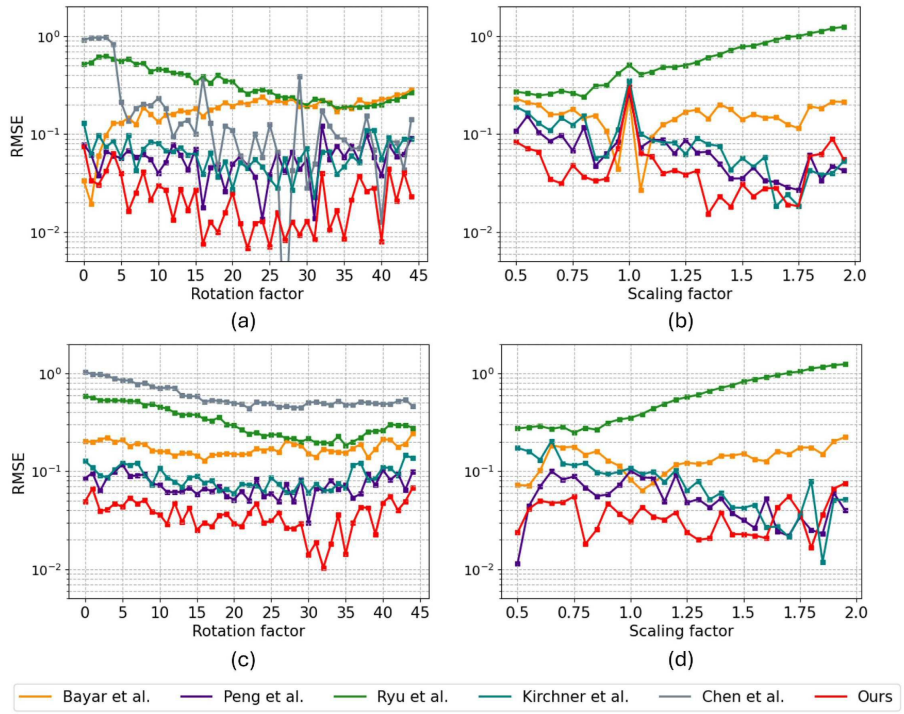
도면3


Original
Test set R
Test set S
Test set M

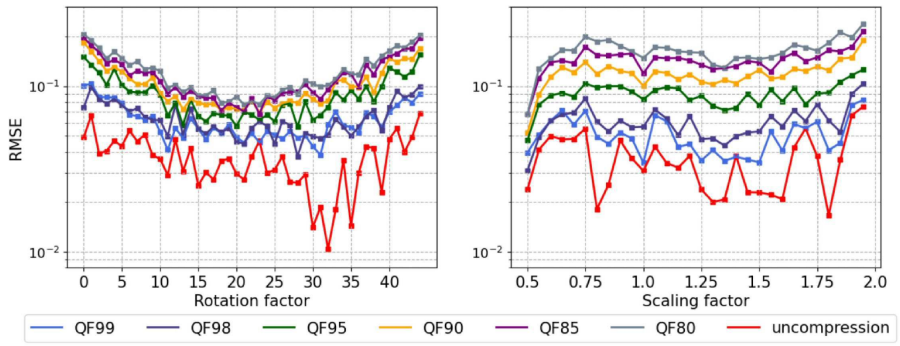
도면4


(a)
(b)
(c)
(d)

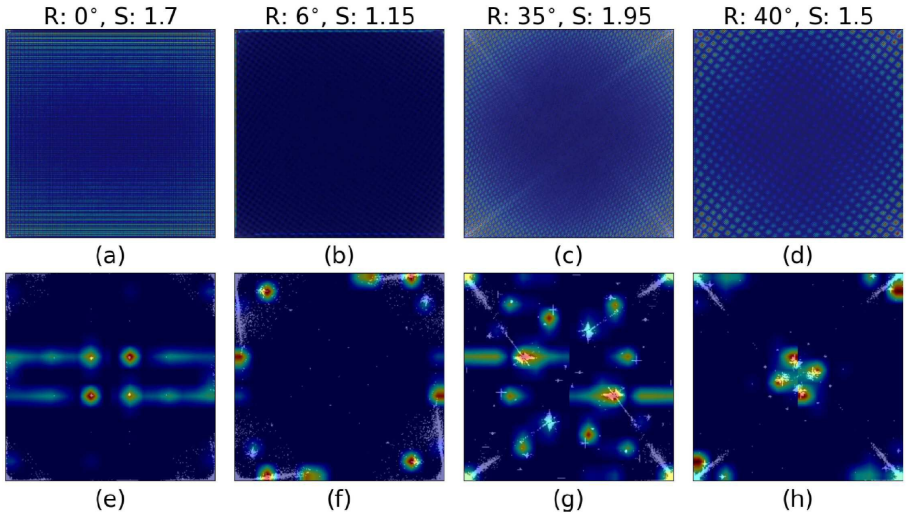
도면5


RMSE
Rotation factor
Scaling factor
(a)
(b)
(c)
(d)
Bayar et al.
Peng et al.
Ryu et al.
Kirchner et al.
Chen et al.
Ours

도면6



도면7



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 5

【변경전】

제1항에 있어서,

상기 FPE(Frequency Positional Encoding) 레이어는,

상기 타겟 이미지에 대하여 1차 이상의 미분이 수행된 잔차(u'_h)에 대하여 2D FFT(Fast Fourier Transform)를 수행하여 주파수 영역으로 변환된 이미지에 대한 함수 값인 $U_h(u, v)$를 획득하고,

상기 $U_h(u, v)$의 크기 성분을 일정 크기의 패치($x \in R^{\times 2}$)로 나누어 이하 수학식에 따른 시퀀스 z_0를 출력하고,

$$E \in \mathbb{R}^{p^2 \times d},\ E_{pos} \in \mathbb{R}^{n \times d}$$

$$Z_0 = [x_p^1 E; x_p^2 E; \ldots; x_p^n E] + E_{pos}$$

상기 n은 패치의 결과 수이고, 상기 p는 패치의 크기이고, 상기 E_{pos}는 위치 인코딩에 해당하는, 전자 장치.

【변경후】

제1항에 있어서,

상기 FPE(Frequency Positional Encoding) 레이어는,

상기 타겟 이미지에 대하여 1차 이상의 미분이 수행된 잔차(u'_h)에 대하여 2D FFT(Fast Fourier Transform)를 수행하여 주파수 영역으로 변환된 이미지에 대한 함수 값인 $U_h(u, v)$를 획득하고,

상기 $U_h(u, v)$의 크기 성분을 일정 크기의 패치($x \in R^{\times 2}$)로 나누어 이하 수학식에 따른 시퀀스 z_0를 출력하고,

$$E \in \mathbb{R}^{p^2 \times d},\ E_{pos} \in \mathbb{R}^{n \times d}$$

$$Z_0 = [x_p^1 E; x_p^2 E; \ldots; x_p^n E] + E_{pos}$$

상기 n은 패치의 결과 수이고, 상기 p는 패치의 크기이고, 상기 E_{pos}는 위치 인코딩에 해당하는, 전자 장치.