(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0109688
(43) 공개일자 2022년08월05일

(51) 국제특허분류(Int. Cl.)
G06T 3/40 (2006.01) *G06N 3/04* (2006.01)
G06N 3/08 (2006.01)
(52) CPC특허분류
G06T 3/4053 (2013.01)
G06N 3/0454 (2013.01)
(21) 출원번호 10-2021-0012998
(22) 출원일자 2021년01월29일
심사청구일자 2021년01월29일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
강석주
서울특별시 마포구 백범로 35 서강대학교
서유림
경기도 성남시 수정구 위례광장로 70 103동 904호
(74) 대리인
이철희

전체 청구항 수 : 총 15 항

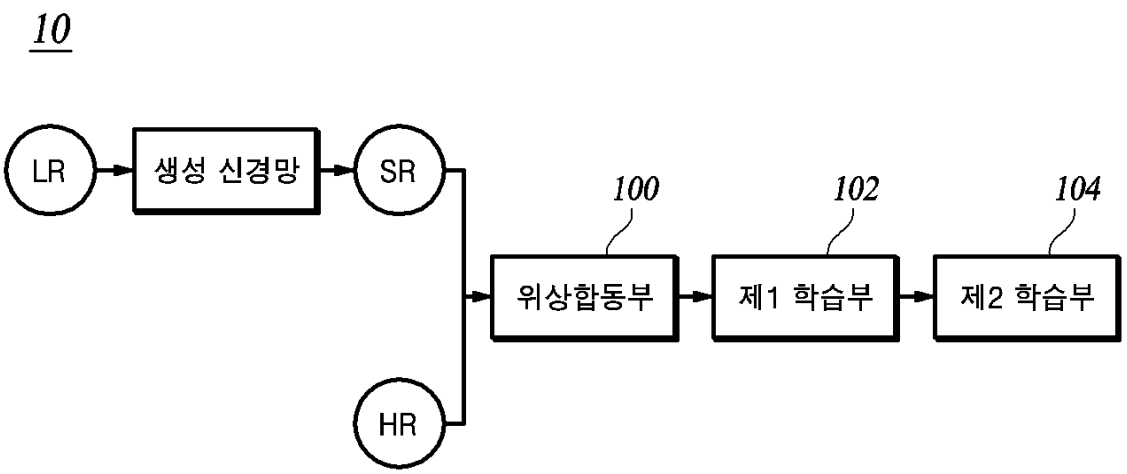
(54) 발명의 명칭 **위상합동을 이용한 초해상화 모델 학습방법 및 그를 위한 초해상화 모델**

(57) 요 약

본 개시는 초해상화 모델의 학습방법 및 그를 위한 초해상화 모델을 제공한다.

본 개시의 일 측면에 의하면, 인공신경망 기반의 초해상화 모델에 있어서, 정답 영상에 대하여 위상합동(Phase Congruency, PC)을 수행하여 주파수 영역(frequency region)을 분할하고, 주파수 영역의 주파수 특성(frequency characteristics)에 따라 맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이 서로 다르게 결정되도록 초해상화 모델을 학습시키는 초해상화 모델의 학습방법 및 그를 위한 초해상화 모델을 제공한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/08 (2013.01)
G06T 3/4046 (2013.01)
G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711119713
과제번호 2020M3H4A1A02084899
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 나노미래소재원천기술개발사업
연구과제명 2축 신축성 AMLED용 TFT 및 LED 집적 기술
기 여 율 1/2
과제수행기관명 경희대학교 산학협력단
연구기간 2020.07.01 ~ 2024.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711116161
과제번호 2018-0-01421-003
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 대학ICT연구센터지원사업
연구과제명 인공지능 서비스 실현을 위한 지능형 반도체 설계 핵심기술 개발
기 여 율 1/2
과제수행기관명 서강대학교 산학협력단
연구기간 2020.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

인공신경망 기반의 초해상화 모델을 학습시키는 방법에 있어서,

정답 영상에 위상합동(Phase Congruency, PC)을 수행하여 주파수 영역(frequency region)을 분할하는 과정; 및

맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이, 상기 주파수 영역의 주파수 특성(frequency characteristics)에 따라 서로 다르게 결정되도록 상기 초해상화 모델을 학습시키는 과정

을 포함하는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 2

제1항에 있어서,

상기 주파수 영역을 분할하는 과정은,

상기 정답 영상을 고주파수 성분(high frequency component)으로 구성된 고주파수 영역(high frequency region)과 저주파수 성분으로 구성된 저주파수 영역(low frequency region)으로 분할하는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 3

제2항에 있어서,

상기 주파수 영역을 분할하는 과정은,

상기 위상합동을 수행하여 생성되는 PC 맵(PC map)의 각 성분을 기 설정된 임계치를 기준으로 1 또는 0으로 변환하여 고주파수용 마스크(high frequency mask) 또는 저주파수용 마스크(low frequency mask)를 생성함으로써 수행되는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 4

제3항에 있어서,

상기 고주파수 영역은, 상기 정답 영상과 상기 고주파수용 마스크 간에 원소별 곱(element-wise multiplication)을 수행함으로써 생성되고,

상기 저주파수 영역은, 상기 정답 영상과 상기 저주파수용 마스크 간에 원소별 곱을 수행함으로써 생성되는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 5

제4항에 있어서,

상기 고주파수용 마스크 또는 상기 저주파수용 마스크에는 바운더리 딜레이션(boundary dilation)이 수행되는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 6

제1항에 있어서,

상기 정답 영상에 상기 위상합동을 수행하여 생성되는 제1 PC 맵과, 학습 전 초해상화 모델로부터 출력한 출력 영상에 상기 위상합동을 수행하여 생성되는 제2 PC 맵 간의 차를 최소화하도록 상기 초해상화 모델을 학습시키는 과정

을 더 포함하는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 7

제1항에 있어서,

상기 인공신경망을 학습시키는 과정은,

상기 주파수 특성이 고주파수에 해당하는 경우 상기 맥락 손실의 비율이 상기 주파수 특성이 저주파수에 해당하는 경우 대비 상대적으로 더 높게 결정되도록 상기 초해상화 모델을 학습시키는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 8

제1항에 있어서,

상기 인공신경망을 학습시키는 과정은,

상기 주파수 특성이 저주파수에 해당하는 경우 상기 인지 손실의 비율이 상기 주파수 특성이 고주파수에 해당하는 경우 대비 상대적으로 더 높게 결정되도록 상기 초해상화 모델을 학습시키는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 9

제1항에 있어서,

상기 맥락 손실의 비율은, 상기 인지 손실의 비율이 증가하면 감소하고, 상기 인지 손실의 비율이 감소하면 증가하는 것을 특징으로 하는 초해상화 모델 학습방법.

청구항 10

제1항 내지 제9항 중 어느 한 항에 따른 초해상화 모델 학습방법이 포함하는 각 과정을 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록매체에 저장된 컴퓨터 프로그램.

청구항 11

비일시적 기록매체에 저장된 인공신경망 기반의 초해상화 모델에 있어서,

정답 영상에 위상합동(Phase Congruency, PC)을 수행하여 주파수 영역(frequency region)을 분할하는 위상합동부; 및

맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이, 상기 주파수 영역의 주파수 특성(frequency characteristics)에 따라 서로 다르게 결정되도록 학습을 수행하는 제1 학습부

를 포함하는 것을 특징으로 하는 초해상화 모델.

청구항 12

제11항에 있어서,

상기 위상합동부는,

상기 위상합동을 수행하여 생성되는 PC 맵(PC map)의 각 성분을 기 설정된 임계치를 기준으로 1 또는 0으로 변환하여 고주파수용 마스크(high frequency mask) 또는 저주파수용 마스크(low frequency mask)를 생성하여, 상기 고주파수용 마스크 또는 상기 저주파수용 마스크를 이용하여 상기 주파수 영역을 분할하는 것을 특징으로 하는 초해상화 모델.

청구항 13

제12항에 있어서,

상기 고주파수용 마스크 또는 상기 저주파수용 마스크에는 바운더리 딜레이션(boundary dilation)이 수행되는 것을 특징으로 하는 초해상화 모델.

청구항 14

제11항에 있어서,

상기 제1 학습부는,

상기 주파수 특성이 고주파수에 해당하는 경우 상기 맥락 손실의 비율이 상기 주파수 특성이 저주파수에 해당하는 경우 대비 상대적으로 더 높게 결정되도록 학습을 수행하고,

상기 주파수 특성이 저주파수에 해당하는 경우 상기 인지 손실의 비율이 상기 주파수 특성이 고주파수에 해당하는 경우 대비 상대적으로 더 높게 결정되도록 학습을 수행하는 것을 특징으로 하는 초해상화 모델.

청구항 15

제11항에 있어서,

상기 정답 영상에 상기 위상합동을 수행하여 생성되는 제1 PC 맵과, 학습 전 초해상화 모델로부터 출력한 출력 영상에 상기 위상합동을 수행하여 생성되는 제2 PC 맵 간의 차를 최소화하도록 학습하는 제2 학습부

를 더 포함하는 것을 특징으로 하는 초해상화 모델.

발명의 설명

기 술 분 야

[0001] 본 발명은 위상합동을 이용한 초해상화 모델 학습방법 및 그를 위한 초해상화 모델에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 최근 합성곱 신경망(Convolutional Neural Network, CNN) 기반의 초해상화(super resolution) 기법에 관한 연구가 다양하게 발전하면서, 종래의 로컬 패치(local patch) 기반 또는 사전 학습 기반의 초해상화와 대비하여 높은 수준의 초해상화가 가능해지고 있다.

[0004] 합성곱 신경망 기반의 초해상화 기법으로는, 픽셀들의 MSE(Mean Square Error)를 최소화하는 방향으로 인공신경망(artificial neural network)을 학습시키는 PSNR-Oriented 기법, 인지 손실(perceptual loss)을 더 고려하여 인공신경망을 학습시키는 SRGAN(Super Resolution Generative Adversarial Network) 기법 등이 있다. 인지 손실이란 비특허문헌 1에서 제안된 손실 함수(loss function)로, 입력 영상과 정답 영상을 다른 딥러닝(deep learning) 기반의 영상 분류 신경망에 입력하여 얻은 피쳐맵(feature map) 간의 차를 최소화하도록 필터 파라미터를 학습시킨다.

[0005] PSNR-Oriented 기법은, 초해상화 성능의 메트릭(metric) 중 하나인 PSNR(Peak-Signal-to-Noise Ration) 값은 향상시키나, 출력 영상과 정답 영상 간의 픽셀별 오차가 줄어들도록 인공신경망을 학습시키므로 출력 영상이 과도하게 편평화(smoothing)되는 문제가 있다. 한편, SRGAN 기법은 인지 손실이 고려되도록 인공신경망을 학습시켜 초해상화 성능을 인지적으로 향상시키나, 출력 영상에 구조적 왜곡이나 잡음을 발생시키는 문제가 있다.

[0006] 그에 따라, 비특허문헌 2에서는 출력 영상과 정답 영상 간의 픽셀 분포의 유사도를 최대화하도록 정답 영상의 픽셀 분포를 이용하는 맥락 손실(contextual loss) 기반의 인공신경망 학습 방법이 제안되었다. 그러나, 맥락 손실 기반의 학습은 인지 손실 기반의 학습 대비 출력 영상에서의 디테일한 복원력이 떨어지고, 출력 영상이 과도하게 편평화되는 문제가 있다.

선행기술문헌

비특허문헌

[0007] (비특허문헌 0001) J. Johnson, A. Alahi, and L. Fei-Fei, “losses for real-time style transfer and super-resolution,” in Proceedings of the European Conference on Computer Vision. Springer, 2016, pp. 694-711.

(비특허문헌 0002) R. Mechrez, I. Talmi, F. Shama, and L. Zelnik-Manor, “natural image statistics with the contextual loss,” in Proceedings of the Asian Conference on Computer Vision. Springer, 2018,

pp. 427-443.

(비특허문헌 0003) X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, “Enhanced super-resolution generative adversarial networks,” in Proceedings of the European Conference on Computer Vision, 2018.

발명의 내용

해결하려는 과제

[0008] 본 개시의 일 측면에 의하면, 인공신경망 기반의 초해상화 모델에 있어서, 정답 영상에 대하여 위상합동(Phase Congruency, PC)을 수행하여 주파수 영역(frequency region)을 분할하고, 주파수 영역의 주파수 특성(frequency characteristics)에 따라 맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이 서로 다르게 결정되도록 초해상화 모델을 학습시키는 초해상화 모델의 학습방법 및 그를 위한 초해상화 모델을 제공하는 데 주된 목적이 있다.

과제의 해결 수단

[0009] 본 개시의 일 측면에 의하면, 인공신경망 기반의 초해상화 모델을 학습시키는 방법에 있어서, 정답 영상에 위상합동(Phase Congruency, PC)을 수행하여 주파수 영역(frequency region)을 분할하는 과정; 및 맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이, 상기 주파수 영역의 주파수 특성(frequency characteristics)에 따라 서로 다르게 결정되도록 상기 초해상화 모델을 학습시키는 과정을 포함하는 것을 특징으로 하는 초해상화 모델 학습방법을 제공한다.

[0010] 본 개시의 다른 측면에 의하면, 전술한 초해상화 모델 학습방법에 있어서, 상기 주파수 영역을 분할하는 과정은, 상기 위상합동을 수행하여 생성되는 PC 맵(PC map)의 각 성분을 기 설정된 임계치를 기준으로 1 또는 0으로 변환하여 고주파수용 마스크(high frequency mask) 또는 저주파수용 마스크(low frequency mask)를 생성함으로써 수행되는 것을 특징으로 하는 초해상화 모델 학습방법을 제공한다.

[0011] 본 개시의 다른 측면에 의하면, 전술한 초해상화 모델 학습방법에 있어서, 상기 정답 영상에 상기 위상합동을 수행하여 생성되는 제1 PC 맵과, 학습 전 초해상화 모델로부터 출력한 출력 영상에 상기 위상합동을 수행하여 생성되는 제2 PC 맵 간의 차를 최소화하도록 상기 초해상화 모델을 학습시키는 과정을 더 포함하는 것을 특징으로 하는 초해상화 모델 학습방법을 제공한다.

[0012] 본 개시의 다른 측면에 의하면, 전술한 초해상화 모델 학습방법에 있어서, 상기 인공신경망을 학습시키는 과정은, 상기 주파수 특성이 고주파수에 해당하는 경우 상기 맥락 손실의 비율이 상기 주파수 특성이 저주파수에 해당하는 경우 대비 상대적으로 더 높게 결정되도록 상기 초해상화 모델을 학습시키는 것을 특징으로 하는 초해상화 모델 학습방법을 제공한다.

[0013] 본 개시의 또 다른 측면에 의하면, 비일시적 기록매체에 저장된 인공신경망 기반의 초해상화 모델에 있어서, 정답 영상에 위상합동(Phase Congruency, PC)을 수행하여 주파수 영역(frequency region)을 분할하는 위상합동부; 및 맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이, 상기 주파수 영역의 주파수 특성(frequency characteristics)에 따라 서로 다르게 결정되도록 학습을 수행하는 제1 학습부를 포함하는 것을 특징으로 하는 초해상화 모델을 제공한다.

발명의 효과

[0014] 본 개시의 일 측면에 의하면, 입력된 영상을 주파수 영역에 따라 분할하고, 분할된 각 영역의 주파수 특성에 따라 손실 함수에 차이가 있도록 적용하도록 초해상화 모델을 학습시킴으로써, 출력 영상의 고주파수 영역에서 발생하는 구조적 왜곡과 잡음을 방지하고, 저주파수 영역에서 발생하는 복원력 감소를 방지하여 종래 기술 대비 선명하게 복원된 영상을 얻는 효과가 있다.

[0015] 본 개시의 다른 측면에 의하면, 위상합동을 이용하여 생성된 주파수별 마스크를 기초로 주파수 영역을 분할함으로써, 입력된 영상의 밝기(brightness)와 대비(contrast)의 영향을 받지 않고도 영상의 각 영역을 주파수 성분별로 분할할 수 있는 효과가 있다.

[0016] 본 개시의 다른 측면에 의하면, 위상합동을 수행하여 생성되는 제1 PC 맵과, 학습 전 초해상화 모델로부터 출력한 출력 영상에 위상합동을 수행하여 생성되는 제2 PC 맵 간의 차를 최소화하도록 초해상화 모델을 더 학습시킴으로써, 입력된 영상의 각 영역을 보다 정확하게 주파수 성분별로 분할하여 분할된 각 영역에 적절한 손실 함수를 적용할 수 있는 효과가 있다.

[0017] 따라서, 본 개시의 여러 측면에 따른 초해상화 모델 학습방법 및 초해상화 모델을 이용하는 경우, 종래의 초해상화 기법을 이용하는 경우 대비 고품질이면서 정답 영상과의 인지적 유사도가 높은 고해상도의 출력 영상을 획득하는 효과가 있다.

도면의 간단한 설명

[0018] 도 1은 본 개시의 일 실시예에 따른 초해상화 모델을 나타내는 개념도이다.

도 2a 및 도 2b는 본 개시의 일 실시예에 따라 영상의 주파수 영역을 분할하는 방법을 나타내는 개념도이다.

도 3은 ESRGAN에 본 개시의 일 실시예에 따른 초해상화 모델을 적용한 예시도이다.

도 4는 본 개시의 일 실시예에 따른 초해상화 모델 학습방법의 순서도이다.

도 5a 내지 도 5c는 본 개시의 일 실시예에 따른 초해상화 모델의 인지적 성능을 평가한 결과 영상이다.

발명을 실시하기 위한 구체적인 내용

[0019] 이하, 본 개시의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다. 각 도면의 구성요소들에 열람부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 개시를 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 개시의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0020] 또한, 본 개시의 구성 요소를 설명하는 데 있어서, 제2, 제1 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 '포함', '구비'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 '…부', '모듈' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0021] 첨부된 도면과 함께 이하에 개시될 상세한 설명은 본 개시의 예시적인 실시형태를 설명하고자 하는 것이며, 본 개시가 실시될 수 있는 유일한 실시형태를 나타내고자 하는 것이 아니다.

[0022] 본 개시에서의 영상(映像, image)은 정지 영상(still image)과 동영상(video)을 모두 포함한다.

[0023] 본 개시에서는 초해상화(super resolution) 기법으로서, 위상합동(Phase Congruency, PC)을 이용하여 정답 영상(ground-truth image)으로부터 주파수 성분을 추출하고, 주파수 성분을 기초로 마스크(mask)를 생성하여 정답 영상의 주파수 영역에 맞춰 학습시키기 위한, 비일시적 기록매체에 저장된 인공신경망(artificial neural network) 기반의 초해상화 모델 및 초해상화 모델을 학습시키기 위한 방법을 제시한다. 본 개시에서의 인공신경망은 심층 신경망(Deep Neural Network, DNN)으로서, 합성곱 신경망(Convolutional Neural Network, CNN)을 포함함이 바람직하나, 이에 한하지 않는다.

[0024] 위상합동이란, 입력 신호를 진폭(amplitude)과 위상(phase)을 가진 여러 코사인 곡선의 합으로 분해할 때, 각 코사인 곡선의 진폭을 크기로 하고 위상을 방향으로 하는 벡터들의 총합(로컬 에너지)을 각 코사인 곡선의 진폭의 총합으로 나누는 연산 방법이다. 각 벡터의 방향이 유사한 경우 위상합동을 수행한 값인 PC 값은 1에 가까워진다. 반면, 각 벡터의 방향이 서로 다르거나 서로 반대 방향인 경우 PC 값은 0에 가까워진다.

[0025] 2차원 이미지 신호에 위상합동을 적용하는 경우, 2차원 PC 값인 PC 맵을 얻는다. 이때, PC 맵의 PC 값이 1에 가까울수록 위상합동의 정도가 큰 것으로, 로컬 에너지가 강항 부분임을 의미한다. 로컬 에너지가 강한 부분은 2차원 이미지의 픽셀 영역에서 에지(edge) 또는 코너(corner)로 나타난다.

[0026] 정답 영상이란, 저해상도 영상(low resolution image)에 초해상화를 수행하는 경우 초해상화가 잘 수행되었는지를 판단할 수 있는 GT(Ground-Truth)이다.

[0027] 도 1은 본 개시의 일 실시예에 따른 초해상화 모델을 나타내는 개념도이다.

[0028] 본 개시의 일 실시예에 따른 초해상화 모델(10)은 위상합동부(phase congruency unit, 100), 제1 학습부(first learning unit, 102) 및 제2 학습부(second learning unit, 104)를 전부 또는 일부 포함한다. 도 1에 도시된 초해상화 모델(10)은 본 개시의 일 실시예에 따른 것으로서, 도 1에 도시된 모든 구성이 필수 구성요소는 아니며, 다른 실시예에서 일부 구성이 추가, 변경 또는 삭제될 수 있다.

[0029] 도 1은 입력 영상(도 1의 LR)을 초해상화한 출력 영상(도 1의 SR)을 생성하는 생성 신경망(generative neural network)에 있어, 초해상화 모델(10)을 생성 신경망과 연결된 판별 신경망(discriminative neural network)으로서 도시하나, 이는 설명의 편의를 위한 것으로 다른 실시예에서 초해상화 모델은 하나의 신경망으로서 구현될 수 있다.

[0030] 위상합동부(100)는 정답 영상(도 1의 HR)에 위상합동을 수행하여 주파수 영역(frequency region)을 분할한다. 위상합동부(100)는 제2 학습부(104)가 위상합동 손실(PC-loss)을 최소화하도록 초해상화 모델(10)을 학습시키기 위하여 생성 신경망이 출력한 출력 영상에 위상합동을 더 수행하여 출력 영상의 주파수 영역을 더 분할할 수 있다.

[0031] 위상합동부(100)는 주파수 영역을 주파수 성분(frequency component)의 주파수 특성(frequency characteristics)에 따라 분할한다. 예컨대, 위상합동부(100)는 입력된 영상의 주파수 영역을 고주파수 성분(high frequency component)으로 구성된 고주파수 영역(high frequency region)과 저주파수 성분으로 구성된 저주파수 영역(low frequency region)으로 분할할 수 있다.

[0032] 이러한 주파수 영역의 분할은, 입력된 영상에 위상합동을 수행하여 얻은 PC 맵의 각 성분을 기 설정된 임계치를 기준으로 1 또는 0으로 변환함으로써 각 주파수 특성에 해당하는 마스크를 생성함으로써 수행될 수 있다. 예컨대, 위상합동부(100)는 PC 맵의 성분이 기 설정된 임계치 이상이면 1, 미만이면 0으로 변환시켜 고주파수용 마스크(high frequency mask)를 생성할 수 있다. 또는, PC 맵의 성분이 기 설정된 임계치 이하이면 1, 초과이면 0으로 변환시켜 저주파수용 마스크(low frequency mask)를 생성할 수 있다. 주파수 영역이 두 가지 유형의 영역으로만 분할되는 경우, 어느 한 마스크의 0과 1을 반전시킴으로써 다른 한 마스크를 생성할 수 있다.

[0033] 위상합동부(100)는 입력된 영상과 각 마스크 간에 원소별 곱(element-wise multiplication)을 수행함으로써 해당 마스크에 대응하는 주파수 영역을 얻는다. 예컨대, 위상합동부(100)는 입력된 영상과 고주파수용 마스크 간에 원소별 곱을 수행하여 고주파수 영역을 획득할 수 있다. 위상합동부(100)는 입력된 영상과 저주파수용 마스크간에 원소별 곱을 수행하여 저주파수 영역을 획득할 수 있다.

[0034] 입력된 영상을 고주파수 영역과 저주파수 영역으로 분할하는 구체적인 방법은 도 2a 및 도 2b에서 후술한다.

[0035] 제1 학습부(102)는 출력 영상과 정답 영역 간 적용되는 손실 함수로서, 맥락 손실(contextual loss)의 비율과 인지 손실(perceptual loss)의 비율이 주파수 영역의 각 주파수 특성에 따라 결정되도록 학습을 수행한다. 이는 초해상화 모델에 맥락 손실의 손실 함수를 적용하는 경우와 인지 손실의 손실 함수를 적용하는 경우 각각에 있어, 영상 내에 초해상화 성능이 우수한 영역이 서로 다르게 나타나기 때문이다. 예컨대, 인지 손실을 최소화하는 손실 함수를 적용하는 경우 영상의 디테일한 부분을 복원시키는 데 우수한 성능을 보이나, 선(line)이나 에지(edge)와 같은 고주파수 성분이 빈번한 영역에서는 구조적 왜곡이나 잡음이 발생하는 문제가 있다.

[0036] 이에 반해, 맥락 손실을 최소화하는 손실 함수를 적용하는 경우 영상의 선이나 에지 부분에서의 구조적 왜곡이나 잡음이 발생하는 문제는 해결되지만, 영상의 디테일한 부분의 복원 성능이 떨어지는 문제가 있다. 복원 성능이 떨어지는 문제는 픽셀 간의 변화가 적은, 즉 영상의 저주파수 성분이 빈번한 영역에서 주로 나타난다. 따라서, 제1 학습부(102)는 영상의 고주파수 영역에서는 맥락 손실이 적용되는 비율이 높아지도록, 저주파수 영역에서는 인지 손실이 적용되는 비율이 높아지도록 학습을 수행한다. 구체적으로, 제1 학습부(102)에는 영상의 고주파수 영역에서는 맥락 손실의 비율이 저주파수 영역에 적용되는 맥락 손실의 비율 대비 상대적으로 더 높아지도록 손실 함수를 구성하고, 영상의 저주파수 영역에서는 인지 손실의 비율이 고주파수 영역에 적용되는 인지 손실의 비율 대비 상대적으로 더 높아지도록 손실 함수를 구성하는 상호 손실(mutual loss)의 손실 함수가 적용되도록 학습할 수 있다.

[0037] 이로써, 초해상화 모델(10)의 출력 영상의 인지적 성능을 향상시키면서, 고주파수 영역에서의 구조적 왜곡이나 잡음을 방지하고 저주파수 영역에서의 디테일을 선명하게 복원하는 효과가 있다.

[0038] 제2 학습부(104)는 정답 영상에 위상합동을 수행하여 생성되는 제1 PC 맵과, 생성 신경망으로부터 출력한 출력 영

상에 위상합동을 수행하여 생성되는 제2 PC 맵 간의 차를 최소화하도록 학습을 수행한다. 제2 학습부(104)의 학습은 제1 학습부(102)의 학습이 수행되기 전에 또는 학습이 수행된 후에 수행될 수 있다.

[0039] 도 2a 및 도 2b는 본 개시의 일 실시예에 따라 영상의 주파수 영역을 분할하는 방법을 나타내는 개념도이다.

[0040] 도 2a에서와 같이, 초해상화 모델은 입력된 영상에 PC 맵 추출기(PC map extractor)를 이용하여 위상합동을 수행하고, 마스크 생성기(mask generator)를 이용하여 획득된 PC 맵의 성분을 기 설정된 기준값을 기초로 1 또는 0으로 변환하여 각각 고주파수용 마스크와 저주파수용 마스크를 생성한다. 저주파수용 마스크는 고주파수용 마스크를 반전시키거나, PC 맵을 기 설정된 기준값을 기초로 0 또는 1로 변환함으로써 획득될 수 있다.

[0041] 초해상화 모델은 각 마스크에 바운더리 딜레이션(boundary dilation)을 더 수행한다. 딜레이션을 수행하는 이유는, 이후 입력된 영상과 마스크간의 곱을 수행할 때 입력된 영상의 해상도 손실 없이 수용 영역(receptive filed)의 크기를 팽창(dilate)시키고, 입력된 영상의 문맥 정보를 보다 잘 추출할 수 있도록 하기 위함이다. 딜레이션은 예컨대, 입력된 영상으로부터 고해상도 성분의 영역과 저해상도 성분의 영역을 획득하는 레이어의 필터에 패딩(padding)을 부가함으로써 수행될 수 있다. 고주파수용 마스크와 저주파수용 마스크 각각에 딜레이션을 수행한 예시는 도 2a의 (a)와 (b)에 나타나 있다.

[0042] 도 2b에서와 같이, 초해상화 모델은 입력된 영상으로부터 고주파수 영역과 저주파수 영역을 각각 획득하기 위하여, 입력된 영상과 고주파수용 마스크 간 및/또는 입력된 영상과 저주파수용 마스크 간에 원소 간 곱을 각각 수행한다.

[0043] 도 2b를 참조하면, 획득된 영상의 고주파수 영역은 이미지의 선이나 에지 부분임을 확인할 수 있고, 저주파수 영역은 그외의 픽셀 간 변화가 적은 부분임을 확인할 수 있다.

[0044] 도 3은 ESRGAN에 본 개시의 일 실시예에 따른 초해상화 모델을 적용한 예시도이다.

[0045] ESRGAN(Enhanced-Super Resolution Generative Adversarial Network)은 SRGAN(Super Resolution Generative Adversarial Network)에 RDDB(Residual-in-Residual Dense Block)을 도입하되, 배치 정규화(batch normalization)가 아닌 잔차 스케일링(residual scaling)을 수행하도록 변형을 가한 것이다(비특허문헌 3 참조). 또한, RaGAN(Relativistic average Generative Adversarial Network)를 사용하여, 판별 신경망이 생성 신경망이 생성한 출력 영상이 현실적인지 여부를 판단하게 한다(비특허문헌 3 참조).

[0046] 도 3을 참조하면, 초해상화 모델은 ESRGAN이 생성한 출력 영상(도 3의 SR)에 정답 영상(도 3의 HR) 간의 손실, 특히 전술한 상호 손실을 연산하여 초해상화 모델을 학습시킨다. 또한, 초해상화 모델은 출력 영상과 정답 영상 각각에 대한 PC 맵을 생성하여 각 PC 맵 간의 손실을 연산하여 초해상화 모델을 학습시킨다. 초해상화 모델은 정답 영상의 각 주파수 영역의 주파수 특성에 따라 맥락 손실과 인지 손실의 비율을 다르게 적용하여 상호 손실을 연산함으로써, 입력된 영상의 고주파수 영역에 발생할 수 있는 구조적 왜곡 또는 잡음과, 저주파수 영역에 발생할 수 있는 복원 성능 감소를 방지할 수 있다.

[0047] 도 4는 본 개시의 일 실시예에 따른 초해상화 모델 학습방법의 순서도이다.

[0048] 입력된 영상에 위상합동을 수행한다(S400).

[0049] 위상합동 수행으로 생성된 PC 맵의 각 성분을 1 또는 0으로 변환하여 고주파수용 마스크 또는 저주파수용 마스크 생성한다(S402).

[0050] 각 마스크와의 원소 간 곱을 수행하여, 입력된 영상을 고주파수 성분으로 구성된 고주파수 영역과 저주파수 성분으로 구성된 저주파수 영역으로 분할한다(S404).

[0051] 각 주파수 영역의 주파수 특성에 따라 맥락 손실의 비율과 인지 손실의 비율이 서로 다르게 결정되도록 초해상화 모델을 학습시킨다(S406). 주파수 특성이 고주파수에 해당하는 경우 맥락 손실의 비율을 상대적으로 높여 인지 손실 기반 학습에 의한 잡음을 방지한다. 주파수 특성이 저주파수에 해당하는 경우 인지 손실의 비율을 상대적으로 높여 맥락 손실 기반 학습에 의한 디테일한 부분이 편평화되는 문제를 방지한다.

[0052] 정답 영상의 PC 맵과 출력 영상의 PC 맵 간의 차를 최소화하도록 초해상화 모델을 더 학습시킨다(S408). S408 단계의 학습으로써 초해상화 모델의 위상합동 연산 성능이 향상된다.

[0053] 도 4에서는 과정 각 과정을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 본 개시의 일 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것이다. 다시 말해, 본 개시의 일 실시예가 속하는 기술 분야에서 통상의

지식을 가진 자라면 본 개시의 일 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 4에 기재된 순서를 변경하여 실행하거나 각 과정 중 하나 이상의 과정을 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 4의 시계열적인 순서로 한정되는 것은 아니다.

[0054] 초해상화 성능을 평가하기 위하여는 각 매트릭(metric)을 이용한 정량적 평가는 물론, 실제 사용자에게 높은 해상도의 고품질 영상으로 판단되는지를 확인할 수 있는 정성적 평가가 모두 요구된다.

[0055] 본 개시에 따라 학습된 초해상화 모델("Proposed")과 바이큐빅 보간법("Bicubic"), SRGAN 기법, EnhancedNet 기법, ESRGAN 기법, RCAN(Residual Channel Attention Network) 기법 및 NatSR(NATural and realistic single image Super Resolution) 기법 각각에 대하여 벤치마크를 수행한 정량적 결과는 표 1에 나타나 있다.

[0056] 한편, 벤치마크를 수행한 정성적 결과는 도 5에 나타나 있다.

[0057] 본 벤치마크에서는, 벤치마크 데이터셋으로서 초해상화 성능 평가에 일반적으로 사용되는 데이터셋인 Set5, Set14, BSD100, Urban100 및 General100을 이용하였다.

[0058] 벤치마크한 메트릭은 PSNR(Peak-Signal-to-Noise Ration), SSIM(Structural Similarity Index Measure) 및 LPIPS(Learned Perceptual Image Patch Similarity)을 이용하였다.

[0059] PSNR이란 출력 영상과 정답 영상 간의 평균 제곱 오차 대비 잡음의 정도에 로그를 취한 메트릭으로, 정답 영상 대비 출력 영상의 손상 정도가 클수록 0에 가까운 값을 가진다. PSNR은 사람이 실제 인지하는 출력 영상의 품질을 제대로 반영하지 못하는 문제가 있다.

[0060] SSIM이란 출력 영상의 구조 정보(휘도(luminance), 대비(contrast), 구조(structure))가 정답 영상의 구조 정보 대비 유지되는 정도를 측정한 메트릭이다. 출력 영상과 정답 영상이 유사할수록 1에 가까운 값을 가진다. SSIM은 출력 영상과 정답 영상 간 인지적 거리(perceptual distance)를 연산하여, 초해상화 성능 평가에 인간의 인지적 판단을 반영시킨 메트릭이다. 그러나 이는 인간의 인지적 판단에 포함된 다양한 뉘앙스(nuances)의 차이를 모두 반영시키지는 못하는 문제가 있다.

[0061] LPIPS는 신경망으로부터 추출되는 특징(feature)을 이용하여 사람이 인지하는 영상과 출력 영상 간 유사도를 측정한 메트릭이다. 이는 출력 영상과 정답 영상이 유사할수록 0에 가까운 값을 가진다. LPIPS는 초해상화 성능 평가에 대한 인간의 인지적 판단에 부합하는지 여부를 객관적으로 평가할 수 있는 메트릭이다.

[0062] 표 1을 참조하면, 본 개시의 초해상화 모델은 LPIPS에서 가장 우수한 성능을 보임을 확인할 수 있다.

표 1

[0063]

Method	Metric	Set5	Set14	BSD100	Urban100	General100
Bicubic	PSNR	28.420	25.990	25.956	23.137	28.006
	SSIM	0.8245	0.7837	0.6672	0.9008	0.8278
	LPIPS	0.3407	0.4384	0.5240	0.4726	0.4384
SRGAN	PSNR	29.833	26.743	25.881	24.419	29.281
	SSIM	0.8567	0.7861	0.6420	0.9314	0.8483
	LPIPS	0.1102	0.2205	0.2551	0.2088	0.1438
EnhanceNet	PSNR	28.838	25.932	25.255	23.623	28.331
	SSIM	0.8409	0.7762	0.6351	0.9317	0.8343
	LPIPS	0.1205	0.1648	0.2081	0.1701	0.1366
ESRGAN	PSNR	30.604	27.352	25.948	25.317	30.352
	SSIM	0.8801	0.8092	0.6448	0.9511	0.8729
	LPIPS	0.0980	0.1673	0.1978	0.1476	0.1110
RCAN	PSNR	32.708	28.819	27.836	27.088	32.038
	SSIM	0.9125	0.8564	0.7446	0.9695	0.9033
	LPIPS	0.1705	0.2759	0.3598	0.1991	0.1674
NatSR	PSNR	30.981	27.418	26.443	25.460	30.342
	SSIM	0.8796	0.8120	0.6826	0.9503	0.8718
	LPIPS	0.0943	0.1765	0.2115	0.1502	0.1117
Prprosed	PSNR	30.806	27.200	25.817	25.250	30.139
	SSIM	0.8774	0.8062	0.6683	0.9528	0.8685
	LPIPS	**0.0661**	**0.1314**	**0.1671**	**0.1250**	**0.0890**

[0064] 도 5a 내지 도 5c는 본 개시의 일 실시예에 따른 초해상화 모델의 인지적 성능을 평가한 출력 영상이다.

[0065] 도 5a 내지 도 5c에서는 동일한 입력 영상 각각에 대하여, 본 개시에 따라 학습된 초해상화 모델이 산출한 출력 영상("Proposed") 및 정답 영상("HR")과, 종래의 초해상화 기법인 바이큐빅 보간법("Bicubic"), SRGAN 기법, EnhancedNet 기법, ESRGAN 기법, RCAN(Residual Channel Attention Network) 기법 및 NatSR(NATural and realistic single image Super Resolution) 기법에 의해 산출된 출력 영상이 나타나 있다.

[0066] 도 5a 내지 도 5c를 참조하면, 저해상도의 입력 영상(도 5a 내지 도 5c의 상자표시 영역)에 대한 정답 영상에 가장 부합한 출력 영상은 본 개시에 따라 학습된 초해상화 모델이 산출한 영상임을 확인할 수 있다. 도 5a 내지 도 5c를 참조하면, 선이 있는 영역에 대하여 바이큐빅 보간법, SRGAN 기법, EnhancedNet 기법, ESRGAN 기법 및 NatSR 기법에 의한 출력 영상에는 모두 구조적 왜곡이 발생하였음을 알 수 있다. 또한 바이큐빅 보간법과 RCAN의 경우 디테일한 부분의 복원력이 떨어짐을 알 수 있다. 반면, 본 개시에 따라 학습된 초해상화 모델에 의한 출력 영상에서는 구조적 왜곡의 발생이 감소하였고, 디테일한 부분 또한 정답 영상의 디테일과 유사하게 복원되었음을 알 수 있다.

[0067] 본 명세서에 설명되는 장치, 부(unit), 과정, 단계 등의 다양한 구현예들은, 디지털 전자 회로, 집적 회로, FPGA(field programmable gate array), ASIC(application specific integrated circuit), 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 및/또는 이들의 조합으로 실현될 수 있다. 이러한 다양한 구현예들은 프로그래밍 가능 시스템상에서 실행 가능한 하나 이상의 컴퓨터 프로그램들로 구현되는 것을 포함할 수 있다. 프로그래밍 가능 시스템은, 저장 시스템, 적어도 하나의 입력 디바이스, 그리고 적어도 하나의 출력 디바이스로부터 데이터 및 명령을 수신하고 이들에게 데이터 및 명령을 전송하도록 결합된 적어도 하나의 프로그래밍 가능 프로세서(이것은 특수 목적 프로세서일 수 있거나 혹은 범용 프로세서일 수 있음)를 포함한다. 컴퓨터 프로그램들(이것은 또한 프로그램들, 소프트웨어, 소프트웨어 애플리케이션들 혹은 코드로서 알려져 있음)은 프로그래밍 가능 프로세서에 대한 명령어들을 포함하며 "컴퓨터가 읽을 수 있는 기록매체"에 저장된다.

[0068] 컴퓨터가 읽을 수 있는 기록매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 이러한 컴퓨터가 읽을 수 있는 기록매체는 ROM, CD-ROM, 자기 테이프, 플로피디스크, 메모리 카드, 하드 디스크, 광자기 디스크, 스토리지 디바이스 등의 비휘발성(non-volatile) 또는 비 일시적인(non-transitory) 매체 또는 데이터 전송 매체(data transmission medium)와 같은 일시적인(transitory) 매체를 더 포함할 수도 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다.

[0069] 본 명세서에 설명되는 시스템들 및 기법들의 다양한 구현예들은, 프로그램가능 컴퓨터에 의하여 구현될 수 있다. 여기서, 컴퓨터는 프로그램가능 프로세서, 데이터 저장 시스템(휘발성 메모리, 비휘발성 메모리, 또는 다른 종류의 저장 시스템이거나 이들의 조합을 포함함) 및 적어도 한 개의 커뮤니케이션 인터페이스를 포함한다. 예컨대, 프로그램가능 컴퓨터는 서버, 네트워크 기기, 셋톱박스, 내장형 장치, 컴퓨터 확장 모듈, 개인용 컴퓨터, 랩톱, PDA(Personal Data Assistant), 클라우드 컴퓨팅 시스템 또는 모바일 장치 중 하나일 수 있다.

[0070] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0071] 10: 초해상화 모델

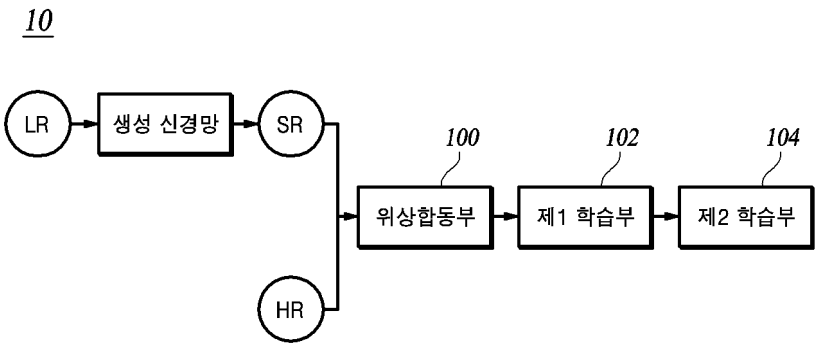
100: 위상합동부

102: 제1 학습부

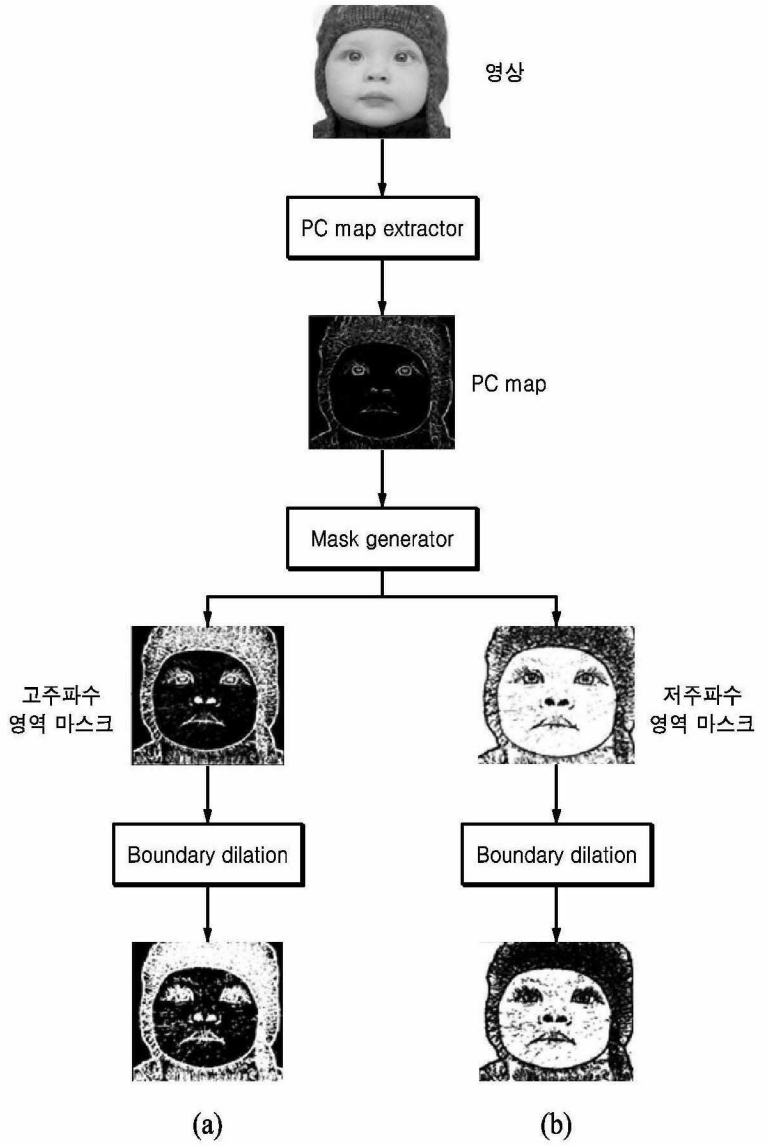
104: 제2 학습부

도면

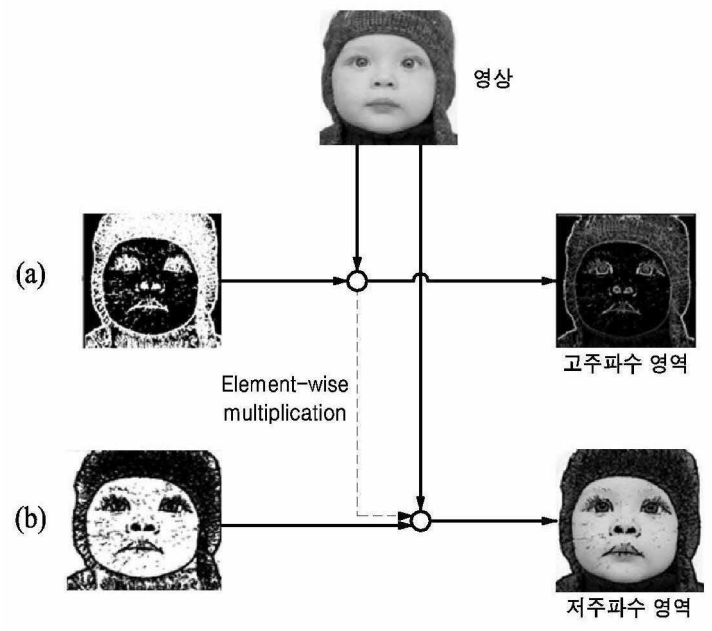
도면1


10
LR
생성 신경망
SR
HR
100
위상합동부
102
제1 학습부
104
제2 학습부

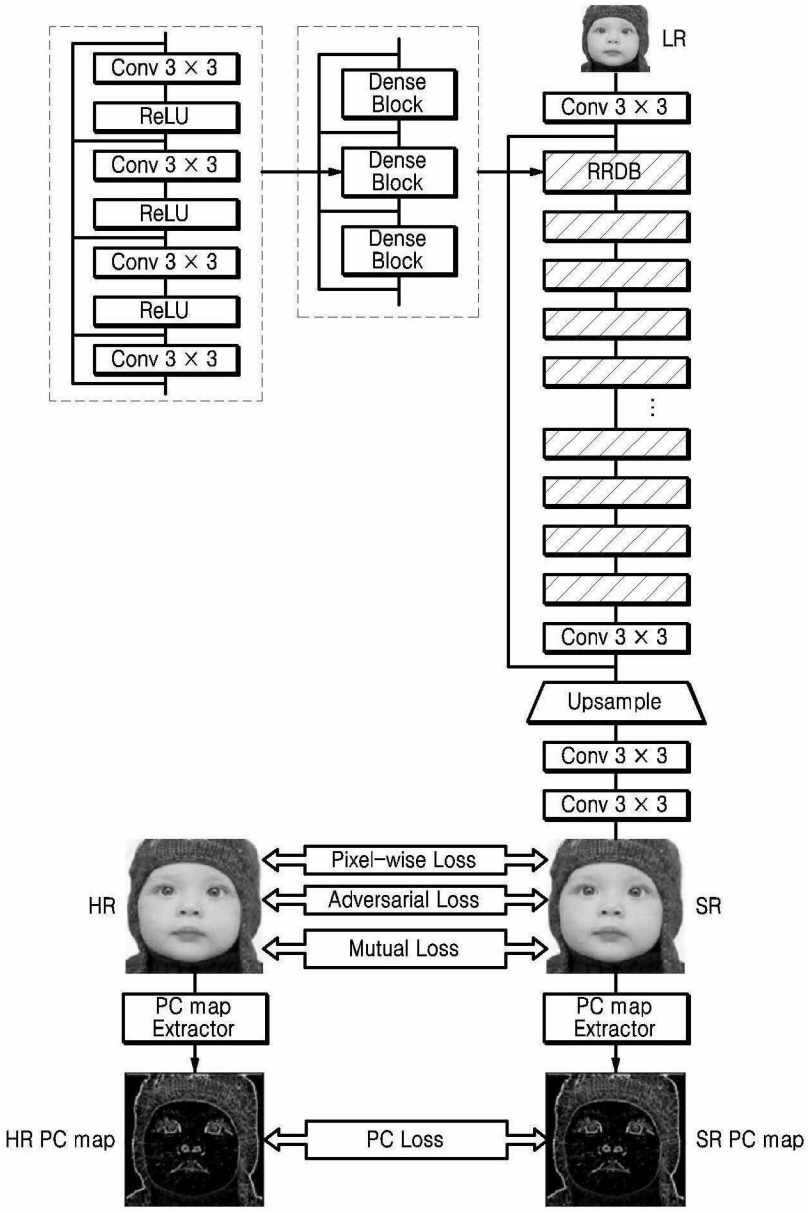
도면2a


영상
PC map extractor
PC map
Mask generator
고주파수
영역 마스크
저주파수
영역 마스크
Boundary dilation
Boundary dilation
(a)
(b)

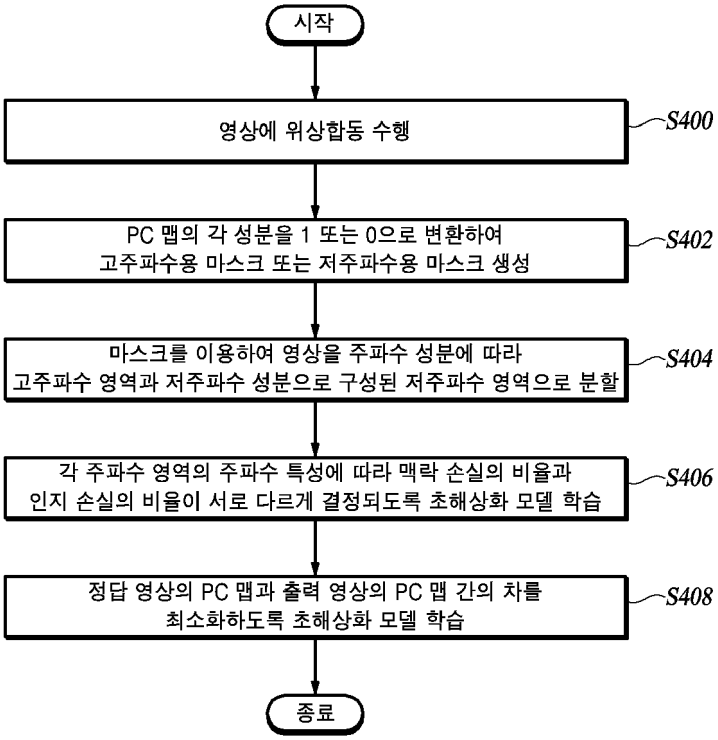
도면2b



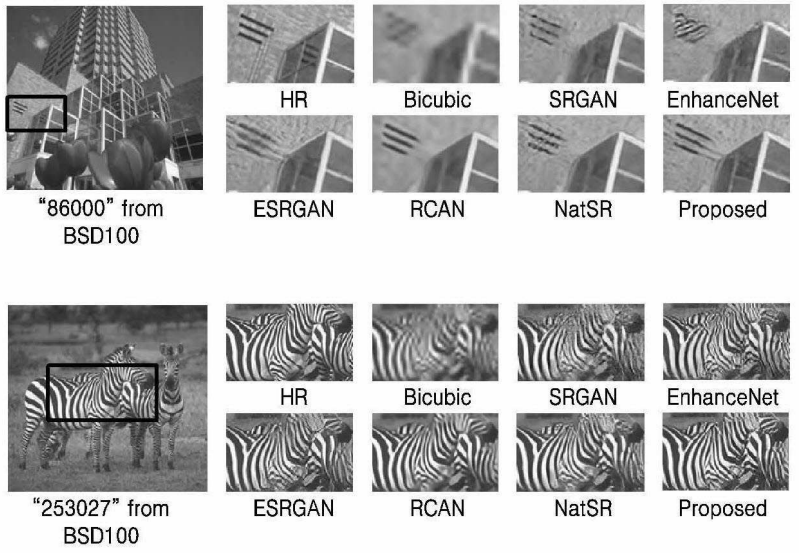
도면3


LR
Conv 3 × 3
ReLU
Conv 3 × 3
ReLU
Conv 3 × 3
ReLU
Conv 3 × 3
Dense Block
Dense Block
Dense Block
RRDB
Conv 3 × 3
Upsample
Conv 3 × 3
Conv 3 × 3
HR
Pixel-wise Loss
Adversarial Loss
Mutual Loss
SR
PC map Extractor
PC map Extractor
HR PC map
PC Loss
SR PC map

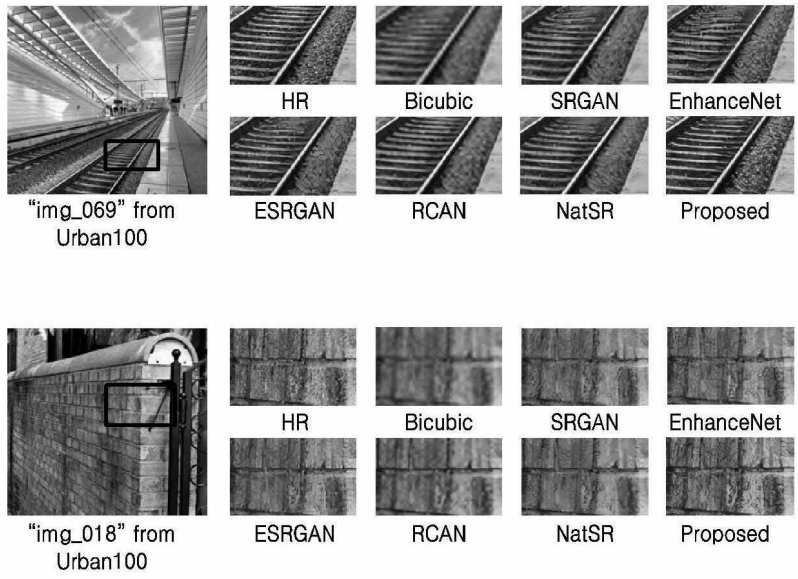
도면4


시작
영상에 위상합동 수행
S400
PC 맵의 각 성분을 1 또는 0으로 변환하여
고주파수용 마스크 또는 저주파수용 마스크 생성
S402
마스크를 이용하여 영상을 주파수 성분에 따라
고주파수 영역과 저주파수 성분으로 구성된 저주파수 영역으로 분할
S404
각 주파수 영역의 주파수 특성에 따라 맥락 손실의 비율과
인지 손실의 비율이 서로 다르게 결정되도록 초해상화 모델 학습
S406
정답 영상의 PC 맵과 출력 영상의 PC 맵 간의 차를
최소화하도록 초해상화 모델 학습
S408
종료

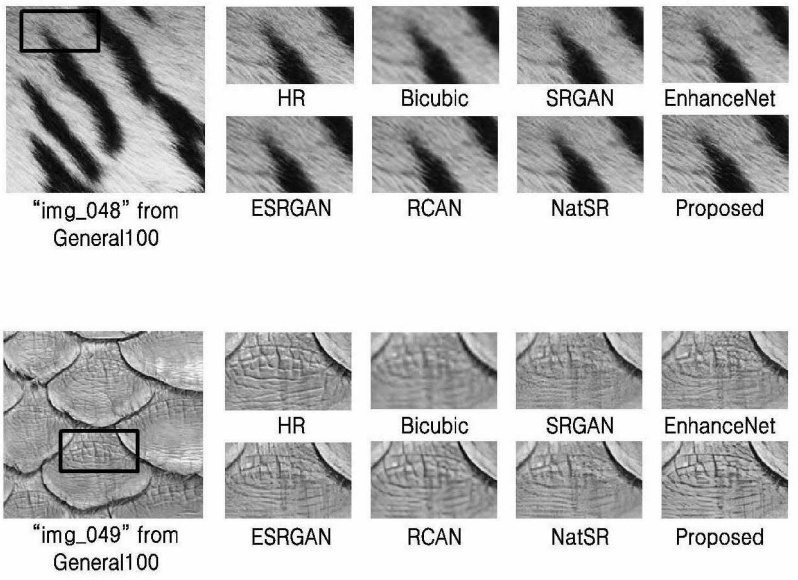
도면5a


HR
Bicubic
SRGAN
EnhanceNet
"86000" from
BSD100
ESRGAN
RCAN
NatSR
Proposed
HR
Bicubic
SRGAN
EnhanceNet
"253027" from
BSD100
ESRGAN
RCAN
NatSR
Proposed

도면5b


HR
Bicubic
SRGAN
EnhanceNet
"img_069" from Urban100
ESRGAN
RCAN
NatSR
Proposed
HR
Bicubic
SRGAN
EnhanceNet
"img_018" from Urban100
ESRGAN
RCAN
NatSR
Proposed

도면5c


HR
Bicubic
SRGAN
EnhanceNet
"img_048" from General100
ESRGAN
RCAN
NatSR
Proposed
HR
Bicubic
SRGAN
EnhanceNet
"img_049" from General100
ESRGAN
RCAN
NatSR
Proposed