

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0055970
(43) 공개일자 2022년05월04일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)
G06T 3/40 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/0454 (2013.01)
(21) 출원번호 10-2020-0140601
(22) 출원일자 2020년10월27일
심사청구일자 2020년10월27일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
함범섭
서울특별시 강남구 압구정로61길 37, 72동 506호
(압구정동, 한양아파트)
이원경
서울특별시 서대문구 연희로 82, B동 604호(연희동, 연희브라운스톤)
(뒷면에 계속)
(74) 대리인
민영준

전체 청구항 수 : 총 19 항

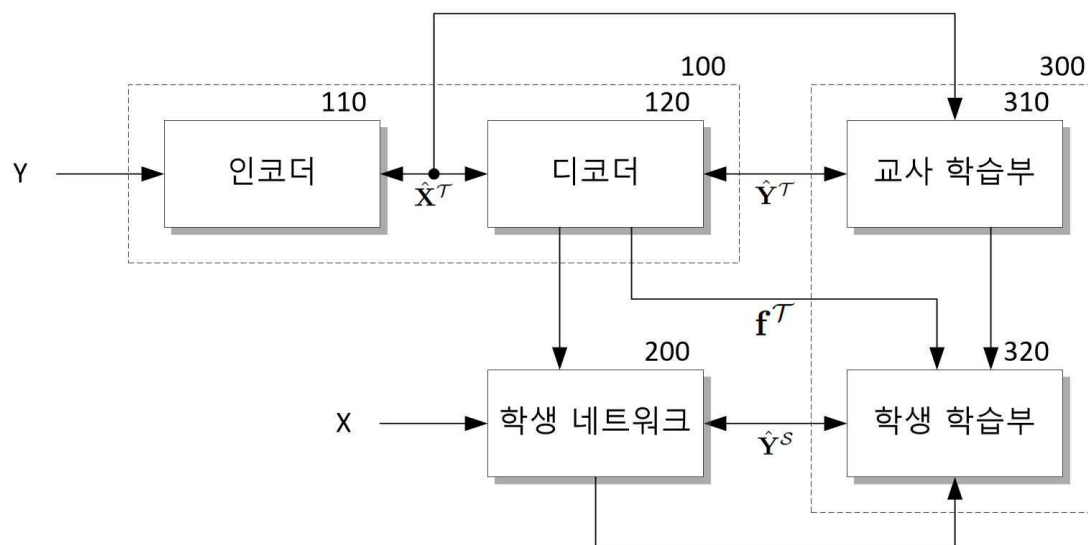
(54) 발명의 명칭 사전 정보 학습 기반 영상 업스케일링 장치 및 방법

(57) 요약

본 발명은 학습 시 결합되는 인코더와 디코더를 포함하는 오토 인코더로 구현되고, 미리 획득된 학습 데이터 셋의 고해상도 영상을 인가받아 학습된 방식에 따라 특징을 추출하여 합축 특징맵을 획득하고, 획득된 합축 특징맵으로부터 학습된 방식에 따라 고해상도 영상을 재구성하여 재구성 영상을 획득하는 교사 네트워크, 교사 네트워

(뒷면에 계속)

대표도 - 도1



크의 디코더와 동일한 구조로 구성되어, 업스케일링 대상이 되는 저해상도의 대상 영상을 인가받아 미리 학습된 방식에 따라 업스케일링하여 고해상도의 업스케일링 영상을 획득하는 학생 네트워크 및 학습 시에 결합되어, 학습 데이터 셋의 고해상도 영상과 대응하는 저해상도 영상, 재구성 영상 및 합축 특징맵을 기반으로 교사 네트워크에 대한 학습을 우선 수행하고, 학습된 교사 네트워크를 기반으로 지식 증류 기법에 따라 학생 네트워크에 학습 데이터 셋의 저해상도 영상을 인가하여 학습을 수행하는 학습부를 포함하여, 사용자 단말에서도 효율적으로 저해상도 영상으로부터 고품질의 고해상도의 영상을 획득할 수 있는 영상 업스케일링 장치 및 방법을 제공할 수 있다.

(52) CPC특허분류

G06T 3/4046 (2013.01)

G06T 3/4053 (2013.01)

(72) 발명자

이중협

서울특별시 서대문구 연희로8길 26(연희동)

김도형

경기도 고양시 덕양구 백양로 8, 1704동 203호(화정동, 옥빛마을17단지아파트)

명세서

청구범위

청구항 1

학습 시 결합되는 인코더와 디코더를 포함하는 오토 인코더로 구현되고, 미리 획득된 학습 데이터 셋의 고해상도 영상을 인가받아 학습된 방식에 따라 특징을 추출하여 함축 특징맵을 획득하고, 획득된 함축 특징맵으로부터 학습된 방식에 따라 고해상도 영상을 재구성하여 재구성 영상을 획득하는 교사 네트워크;

상기 교사 네트워크의 디코더와 동일한 구조로 구성되어, 업스케일링 대상이 되는 저해상도의 대상 영상을 인가받아 미리 학습된 방식에 따라 업스케일링하여 고해상도의 업스케일링 영상을 획득하는 학생 네트워크; 및

학습 시에 결합되어, 상기 학습 데이터 셋의 고해상도 영상과 대응하는 저해상도 영상, 상기 재구성 영상 및 상기 함축 특징맵을 기반으로 상기 교사 네트워크에 대한 학습을 우선 수행하고, 학습된 교사 네트워크를 기반으로 지식 증류 기법에 따라 학생 네트워크에 상기 학습 데이터 셋의 저해상도 영상을 인가하여 학습을 수행하는 학습부를 포함하는 영상 업스케일링 장치.

청구항 2

제1항에 있어서, 상기 인코더는

학습에 의해 업데이트되는 가중치를 기반으로 각각 기지정된 연산을 수행하는 다수의 연산 레이어를 포함하는 인공 신경망으로 구현되어, 학습된 방식에 따라 인가된 학습 데이터 셋의 고해상도 영상의 특징을 추출하여, 상기 저해상도 영상에 대응하는 크기를 갖는 상기 함축 특징맵을 획득하는 영상 업스케일링 장치.

청구항 3

제2항에 있어서, 상기 디코더는

학습에 의해 업데이트되는 가중치를 기반으로 각각 기지정된 연산을 수행하는 다수의 연산 레이어를 포함하는 인공 신경망으로 구현되어, 학습된 방식에 따라 인가된 상기 함축 특징맵을 상기 고해상도 영상에 대응하는 크기를 갖는 상기 재구성 영상으로 재구성하는 영상 업스케일링 장치.

청구항 4

제3항에 있어서, 상기 학습부는

상기 교사 네트워크에 대한 학습 수행 시, 상기 함축 특징맵($\hat{\mathbf{X}}^T$)과 저해상도 영상(X)을 기반으로 모방 손실(L_{im}^T)을 수학식

$$L_{\text{im}}^T = \frac{1}{H'W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} |X_{ij} - \hat{X}_{ij}^T|$$

(여기서 i, j는 픽셀 위치 좌표를 나타내고, H' 과 W' 은 각각 저해상도 영상(X)의 높이와 폭을 나타낸다.)

에 따라 계산하는 영상 업스케일링 장치.

청구항 5

제4항에 있어서, 상기 학습부는

상기 재구성 영상($\hat{\mathbf{Y}}^T$)과 상기 고해상도 영상(Y)을 기반으로 교사 복원 손실(L^T_{recon})을 수학식

$$L^T_{\text{recon}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^T|$$

(여기서 i, j는 픽셀 위치 좌표를 나타내고, H 과 W 은 각각 고해상도 영상(Y)의 높이와 폭을 나타낸다.)

에 따라 계산하는 영상 업스케일링 장치.

청구항 6

제5항에 있어서, 상기 학습부는

계산된 모방 손실(L^T_{im})과 교사 복원 손실(L^T_{recon})로부터 교사 총손실(L^T_{total})을 수학식

$$L^T_{\text{total}} = L^T_{\text{recon}} + \lambda^T L^T_{\text{im}}$$

(여기서 λ^T 는 모방 손실(L^T_{im})과 교사 복원 손실(L^T_{recon})의 비중을 조절하기 위한 파라미터이다.)

에 따라 획득하여, 상기 교사 총손실(L^T_{total})을 상기 교사 네트워크로 역전파하는 영상 업스케일링 장치.

청구항 7

제6항에 있어서, 상기 학습부는

상기 학생 네트워크에 대한 학습 수행 시, 상기 학습 데이터 셋의 저해상도 영상이 인가된 학생 네트워크에서 출력되는 업스케일링 영상($\hat{\mathbf{Y}}^S$)과 고해상도 영상(Y)을 기반으로 수학식

$$L^S_{\text{recon}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^S|$$

에 따라 학생 복원 손실(L^S_{recon})을 계산하는 영상 업스케일링 장치.

청구항 8

제7항에 있어서, 상기 학습부는

상기 학생 네트워크에 대한 추가 학습 수행 시, 상기 학습 데이터 셋의 저해상도 영상이 인가된 학생 네트워크의 다수의 연산 레이어 중 기지정된 하나의 연산 레이어에서 출력되는 중간 특징맵(f^S)과 디코더의 대응하는 위치의 연산 레이어에서 출력되는 중간 특징맵(f^T)을 인가받고, 다변량 라플라스 분포(multivariate Laplace distribution)를 기반으로 하는 중간 특징맵(f^T , f^S) 사이의 파라메트릭 모델($q(f^T; f^S)$)을 이용하여 중간 특징맵(f^T , f^S) 사이의 픽셀 위치에 대한 파라메트릭인 위치맵(μ)과 스케일에 대한 파라메트릭인 스케일맵(b)을 추정하는 영상 업스케일링 장치.

청구항 9

제8항에 있어서, 상기 학습부는

상기 위치맵(μ)과 상기 스케일맵(b) 및 상기 교사 네트워크에서 인가된 중간 특징맵(f^T)을 기반으로 증류 손실(L^S_{distill})을 수학식

$$L_{\text{distill}}^S = \frac{1}{CH'W'} \sum_{i=1}^C \sum_{j=1}^{H'} \sum_{k=1}^{W'} \log b_{ijk} + \frac{|f_{ijk}^T - \mu_{ijk}|}{b_{ijk}}$$

(여기서 C는 중간 특징맵(f^T , f^S)의 채널 수이고, (i,j,k)는 중간 특징맵(f^T , f^S)의 픽셀 위치)

에 따라 계산하는 영상 업스케일링 장치.

청구항 10

제9항에 있어서, 상기 학습부는

상기 학생 복원 손실(L_{recon}^S)과 상기 증류 손실(L_{distill}^S)로부터 학생 총손실(L_{total}^T)을 수학식

$$L_{\text{total}}^S = L_{\text{recon}}^S + \lambda^S L_{\text{distill}}^S$$

에 따라 획득하여, 상기 학생 총손실(L_{total}^S)을 상기 학생 네트워크로 역전파하는 영상 업스케일링 장치.

청구항 11

학생 네트워크와 학습 시에 결합되는 교사 네트워크 및 학습부를 포함하는 영상 업스케일링 장치의 영상 스케일링 방법에 있어서,

상기 학습부가 인코더와 디코더를 포함하는 오토 인코더로 구현되는 상기 교사 네트워크에 미리 획득된 학습 데이터 셋의 고해상도 영상을 입력하고, 학습된 방식에 따라 상기 고해상도 영상의 특징을 추출하여 함축 특징맵이 획득되면, 획득된 함축 특징맵으로부터 학습된 방식에 따라 고해상도 영상을 재구성하여 재구성 영상을 획득하고, 상기 학습 데이터 셋의 고해상도 영상과 대응하는 저해상도 영상, 상기 재구성 영상 및 상기 함축 특징맵을 기반으로 상기 교사 네트워크를 학습시키는 단계;

상기 교사 네트워크에 대한 학습이 완료되면, 학습된 교사 네트워크를 기반으로 지식 증류 기법에 따라 학생 네트워크에 상기 학습 데이터 셋의 저해상도 영상을 인가하여 상기 학생 네트워크를 학습시키는 단계; 및

상기 학생 네트워크에 대한 학습이 완료된 이후, 영상 업스케일링 동작 시에, 학생 네트워크가 업스케일링 대상이 되는 저해상도의 대상 영상을 인가받아 미리 학습된 방식에 따라 업스케일링하여 고해상도의 업스케일링 영상을 획득하는 단계를 포함하는 영상 업스케일링 방법.

청구항 12

제11항에 있어서, 상기 교사 네트워크를 학습시키는 단계는

학습에 의해 업데이트되는 가중치를 기반으로 각각 지정된 연산을 수행하는 다수의 연산 레이어를 포함하는 상기 인코더에 상기 학습 데이터 셋의 고해상도 영상을 입력하여, 상기 저해상도 영상에 대응하는 크기를 갖는 상기 함축 특징맵을 획득하는 단계;

학습에 의해 업데이트되는 가중치를 기반으로 각각 지정된 연산을 수행하는 다수의 연산 레이어를 포함하는 상기 디코더가 상기 함축 특징맵을 인가받아 상기 고해상도 영상에 대응하는 크기를 갖는 상기 재구성 영상을 재구성하는 단계;

상기 학습 데이터 셋의 고해상도 영상과 상기 재구성 영상 사이의 차에 따라 지정된 방식으로 교사 복원 손실을 계산하는 단계;

상기 저해상도 영상과 상기 함축 특징맵 사이의 차에 따라 지정된 방식으로 모방 손실을 계산하는 단계; 및

계산된 상기 교사 복원 손실과 상기 모방 손실을 이용하여 지정된 방식으로 교사 총손실을 획득하여 상기 교사 네트워크로 역전파하는 단계를 포함하는 영상 업스케일링 방법.

청구항 13

제12항에 있어서, 상기 학생 네트워크를 학습시키는 단계는

학생 네트워크에서 저해상도 영상으로부터 상기 업스케일링 영상을 획득하는 과정에서 생성되는 중간맵과 상기 디코더에서 상기 합축 특징맵으로부터 상기 재구성 영상을 획득하는 과정에서 생성되는 중간맵을 획득하여 기 지정된 방식으로 증류 손실을 계산하는 단계;

상기 학습 데이터 셋의 고해상도 영상과 상기 업스케일링 영상 사이의 차에 따라 기지정된 방식으로 학생 복원 손실을 계산하는 단계; 및

계산된 상기 학생 복원 손실과 상기 증류 손실을 이용하여 기지정된 방식으로 학생 총손실을 획득하여 상기 학생 네트워크로 역전파하는 단계를 포함하는 영상 업스케일링 방법.

청구항 14

제13항에 있어서, 상기 교사 복원 손실을 계산하는 단계는

상기 교사 복원 손실(L_{recon}^T)을 수학식

$$L_{\text{recon}}^T = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^T|$$

(여기서 Y 는 고해상도 영상이고, \hat{Y}^T 는 재구성 영상이며, i, j 는 픽셀 위치 좌표를 나타내고, H 과 W 은 각각 고해상도 영상(Y)의 높이와 폭을 나타낸다.)

에 따라 계산하는 영상 업스케일링 방법.

청구항 15

제14항에 있어서, 상기 모방 손실을 계산하는 단계는

상기 모방 손실(L_{im}^T)을 수학식

$$L_{\text{im}}^T = \frac{1}{H'W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} |X_{ij} - \hat{X}_{ij}^T|$$

(여기서 X 는 저해상도 영상이고, \hat{X}^T 는 합축 특징맵이며, i, j 는 픽셀 위치 좌표를 나타내고, H' 과 W' 은 각각 저해상도 영상(X)의 높이와 폭을 나타낸다.)

에 따라 계산하는 영상 업스케일링 방법.

청구항 16

제15항에 있어서, 상기 교사 네트워크로 역전파하는 단계는

상기 교사 총손실(L_{total}^T)을 수학식

$$L_{\text{total}}^T = L_{\text{recon}}^T + \lambda^T L_{\text{im}}^T$$

(여기서 λ^T 는 모방 손실(L_{im}^T)과 교사 복원 손실(L_{recon}^T)의 비중을 조절하기 위한 파라미터이다.)

에 따라 획득하여 역전파하는 영상 업스케일링 방법.

청구항 17

제16항에 있어서, 상기 증류 손실을 계산하는 단계는

상기 학습 데이터 셋의 저해상도 영상이 인가된 학생 네트워크의 다수의 연산 레이어 중 기지정된 하나의 연산

레이어에서 출력되는 중간 특징맵(f^S)과 디코더의 대응하는 위치의 연산 레이어에서 출력되는 중간 특징맵(f^T)을 획득하는 단계;

다변량 라플라스 분포를 기반으로 하는 중간 특징맵(f^T , f^S) 사이의 파라메트릭 모델($q(f^T; f^S)$)을 이용하여 중간 특징맵(f^T , f^S) 사이의 픽셀 위치에 대한 파라메트릭인 위치맵(μ)과 스케일에 대한 파라메트릭인 스케일맵(b)을 추정하는 단계; 및

상기 위치맵(μ)과 상기 스케일맵(b) 및 상기 교사 네트워크에서 인가된 중간 특징맵(f^T)을 기반으로 상기 증류 손실(L^S_{distill})을 수학식

$$L^S_{\text{distill}} = \frac{1}{CH'W'} \sum_{i=1}^C \sum_{j=1}^{H'} \sum_{k=1}^{W'} \log b_{ijk} + \frac{|f^T_{ijk} - \mu_{ijk}|}{b_{ijk}}$$

(여기서 C 는 중간 특징맵(f^T , f^S)의 채널 수이고, (i, j, k) 는 중간 특징맵(f^T , f^S)의 픽셀 위치)

에 따라 계산하는 단계를 포함하는 영상 업스케일링 방법.

청구항 18

제17항에 있어서, 상기 학생 복원 손실을 계산하는 단계는

상기 학생 복원 손실(L^S_{recon})을 수학식

$$L^S_{\text{recon}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}^S_{ij}|$$

(여기서 \hat{Y}^S 는 업스케일링 영상)

에 따라 계산하는 영상 업스케일링 방법.

청구항 19

제18항에 있어서, 상기 학생 네트워크로 역전파하는 단계는

상기 학생 복원 손실(L^S_{recon})과 상기 증류 손실(L^S_{distill})로부터 학생 총손실(L^S_{total})을 수학식

$$L^S_{\text{total}} = L^S_{\text{recon}} + \lambda^S L^S_{\text{distill}}$$

에 따라 획득하여, 상기 학생 총손실(L^S_{total})을 상기 학생 네트워크로 역전파하는 영상 업스케일링 방법.

발명의 설명

기술 분야

[0001] 본 발명은 영상 업스케일링 장치 및 방법에 관한 것으로, 사전 정보와 지식 증류 기법으로 학습되어 영상을 업스케일링 하는 영상 업스케일링 장치 및 방법에 관한 것이다.

배경 기술

[0002] 최근 영상 기술의 발전으로 인해 고해상도 영상에 대한 요구가 증가되고 있으며, 이에 저해상도 영상을 초고해상도 영상으로 변환하는 업스케일링(Upscaling) 기술에 또한 관심을 받고 있다.

[0003] 영상 업스케일링 기술로 다양한 알고리즘이 제안되었으나, 단순한 산술 연산에 기반한 알고리즘보다 인공 신경

망을 이용하여 연산하는 방법인 초해상도(Super-Resolution) 기법이 더 좋은 성능을 나타내는 것으로 알려져 있다.

[0004] 초해상도 기법에서는 컨볼루션(Convolution)연산을 기반으로 하는 CNN(Convolution Neural Network) 알고리즘을 이용하였다. 그러나 CNN은 영상 품질을 향상시키기 위해 많은 메모리와 연산량을 요구할 뿐만 아니라, 업스케일링 속도가 느리다는 문제가 있다.

[0005] 이에 적은 수의 컨볼루션 레이어로 구성되어 단순한 구조를 갖는 SRCNN(Super-Resolution Convolutional Neural Network) 또는 FSRCNN(Fast SRCNN)이 제안되었다. SRCNN이나 FSRCNN은 구조가 복잡하지 않아, TV나 스마트폰과 같은 사용자 단말 수준에서도 빠르게 실행될 수 있으나, 요구되는 고품질의 고해상도 영상을 획득하기 어렵다는 한계가 있다.

[0006] 최근에는 어텐션(attention) 구조나 재귀(recursive) 구조 등의 이용하여 복잡한 네트워크 구조를 갖지만 속도 및 메모리 측면에서 효율적이고 높은 업스케일링 성능을 나타낼 수 있는 기법이 제안되었으나, 복잡한 구조로 인해 여전히 하드웨어로 구현이 어렵다는 문제가 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국 등록 특허 제10-2061935호 (2019.12.26 등록)

발명의 내용

해결하려는 과제

[0008] 본 발명의 목적은 저해상도 영상을 고품질의 고해상도의 영상으로 업스케일링할 수 있는 영상 업스케일링 장치 및 방법을 제공하는데 있다.

[0009] 본 발명의 다른 목적은 사전 정보에 기반하여 학습된 교사 네트워크의 학습 정보를 지식 증류 기법으로 전달받은 학생 네트워크를 이용하여 간단한 구조로 효율적으로 고품질의 고해상도의 영상을 획득할 수 있는 영상 업스케일링 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0010] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 영상 업스케일링 장치는 학습 시 결합되는 인코더와 디코더를 포함하는 오토 인코더로 구현되고, 미리 획득된 학습 데이터 셋의 고해상도 영상을 인가받아 학습된 방식에 따라 특징을 추출하여 합축 특징맵을 획득하고, 획득된 합축 특징맵으로부터 학습된 방식에 따라 고해상도 영상을 재구성하여 재구성 영상을 획득하는 교사 네트워크; 상기 교사 네트워크의 디코더와 동일한 구조로 구성되어, 업스케일링 대상이 되는 저해상도의 대상 영상을 인가받아 미리 학습된 방식에 따라 업스케일링하여 고해상도의 업스케일링 영상을 획득하는 학생 네트워크; 및 학습 시에 결합되어, 상기 학습 데이터 셋의 고해상도 영상과 대응하는 저해상도 영상, 상기 재구성 영상 및 상기 합축 특징맵을 기반으로 상기 교사 네트워크에 대한 학습을 우선 수행하고, 학습된 교사 네트워크를 기반으로 지식 증류 기법에 따라 학생 네트워크에 상기 학습 데이터 셋의 저해상도 영상을 인가하여 학습을 수행하는 학습부를 포함한다.

[0011] 상기 인코더는 학습에 의해 업데이트되는 가중치를 기반으로 각각 기지정된 연산을 수행하는 다수의 연산 레이어를 포함하는 인공 신경망으로 구현되어, 학습된 방식에 따라 인가된 학습 데이터 셋의 고해상도 영상의 특징을 추출하여, 상기 저해상도 영상에 대응하는 크기를 갖는 상기 합축 특징맵을 획득할 수 있다.

[0012] 상기 디코더는 학습에 의해 업데이트되는 가중치를 기반으로 각각 기지정된 연산을 수행하는 다수의 연산 레이어를 포함하는 인공 신경망으로 구현되어, 학습된 방식에 따라 인가된 상기 합축 특징맵을 상기 고해상도 영상에 대응하는 크기를 갖는 상기 재구성 영상으로 재구성할 수 있다.

[0013] 상기 학생 네트워크는 상기 디코더와 동일한 구조를 갖고 지식 증류 기법에 따라 상기 디코더의 다수의 연산 레이어에서 대응하는 연산 레이어의 가중치가 전이되어 초기 설정되는 다수의 연산 레이어를 포함하는 인공 신경망으로 구현될 수 있다.

[0014] 상기 학습부는 상기 교사 네트워크에 대한 학습 수행 시, 상기 함축 특징맵($\hat{\mathbf{X}}^{\mathcal{T}}$)과 저해상도 영상(X)을 기반으로 모방 손실($L_{\text{im}}^{\mathcal{T}}$)을 수학식

$$L_{\text{im}}^{\mathcal{T}} = \frac{1}{H'W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} |X_{ij} - \hat{X}_{ij}^{\mathcal{T}}|$$

[0015] (여기서 i, j는 픽셀 위치 좌표를 나타내고, H' 과 W' 은 각각 저해상도 영상(X)의 높이와 폭을 나타낸다.)에 따라 계산할 수 있다.

[0016]

[0017] 상기 학습부는 상기 재구성 영상($\hat{\mathbf{Y}}^{\mathcal{T}}$)과 상기 고해상도 영상(Y)을 기반으로 교사 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)을 수학식

$$L_{\text{recon}}^{\mathcal{T}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^{\mathcal{T}}|$$

[0018] (여기서 i, j는 픽셀 위치 좌표를 나타내고, H 과 W 은 각각 고해상도 영상(Y)의 높이와 폭을 나타낸다.)에 따라 계산할 수 있다.

[0019]

[0020] 상기 학습부는 계산된 모방 손실($L_{\text{im}}^{\mathcal{T}}$)과 교사 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)로부터 교사 총손실($L_{\text{total}}^{\mathcal{T}}$)을 수학식

$$L_{\text{total}}^{\mathcal{T}} = L_{\text{recon}}^{\mathcal{T}} + \lambda^{\mathcal{T}} L_{\text{im}}^{\mathcal{T}}$$

[0021] (여기서 $\lambda^{\mathcal{T}}$ 는 모방 손실($L_{\text{im}}^{\mathcal{T}}$)과 교사 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)의 비중을 조절하기 위한 파라미터이다.)에 따라 획득하여, 상기 교사 총손실($L_{\text{total}}^{\mathcal{T}}$)을 상기 교사 네트워크로 역전파할 수 있다.

[0022]

[0023] 상기 학습부는 지식 정보가 전이된 학생 네트워크에 대한 추가 학습 수행 시, 상기 학습 데이터 셋의 저해상도 영상이 인가된 학생 네트워크에서 출력되는 업스케일링 영상($\hat{\mathbf{Y}}^{\mathcal{S}}$)과 고해상도 영상(Y)을 기반으로 수학식

$$L_{\text{recon}}^{\mathcal{S}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^{\mathcal{S}}|$$

[0024] 에 따라 학생 복원 손실($L_{\text{recon}}^{\mathcal{S}}$)을 계산할 수 있다.

[0025]

[0026] 상기 학습부는 지식 정보가 전이된 학생 네트워크에 대한 추가 학습 수행 시, 상기 학습 데이터 셋의 저해상도 영상이 인가된 학생 네트워크의 다수의 연산 레이어 중 최종 연산 레이어를 제외한 연산 레이어 중 기지정된 하나의 연산 레이어에서 출력되는 중간 특징맵($f^{\mathcal{S}}$)과 디코더의 대응하는 위치의 연산 레이어에서 출력되는 중간 특징맵($f^{\mathcal{T}}$)을 인가받고, 다변량 라플라스 분포(multivariate Laplace distribution)를 기반으로 하는 중간 특징맵($f^{\mathcal{T}}$, $f^{\mathcal{S}}$) 사이의 파라메트릭 모델($q(f^{\mathcal{T}}; f^{\mathcal{S}})$)을 이용하여 중간 특징맵($f^{\mathcal{T}}$, $f^{\mathcal{S}}$) 사이의 픽셀 위치에 대한 파라메트릭인 위치맵(μ)과 스케일에 대한 파라메트릭인 스케일맵(b)을 추정할 수 있다.

[0027] 상기 학습부는 상기 위치맵(μ)과 상기 스케일맵(b) 및 상기 교사 네트워크에서 인가된 중간 특징맵($f^{\mathcal{T}}$)을 기반으로 증류 손실($L_{\text{distill}}^{\mathcal{S}}$)을 수학식

$$L_{\text{distill}}^{\mathcal{S}} = \frac{1}{CH'W'} \sum_{i=1}^C \sum_{j=1}^{H'} \sum_{k=1}^{W'} \log b_{ijk} + \frac{|f_{ijk}^{\mathcal{T}} - \mu_{ijk}|}{b_{ijk}}$$

[0028] (여기서 C는 중간 특징맵($f^{\mathcal{T}}$, $f^{\mathcal{S}}$)의 채널 수이고, (i,j,k)는 중간 특징맵($f^{\mathcal{T}}$, $f^{\mathcal{S}}$)의 픽셀 위치)에 따라 계산할

[0029]

수 있다.

[0030] 상기 학습부는 상기 학생 복원 손실(L^S_{recon})과 상기 증류 손실(L^S_{distill})로부터 학생 총손실(L^T_{total})을 수학식

$$L^S_{\text{total}} = L^S_{\text{recon}} + \lambda^S L^S_{\text{distill}}$$

[0032] 에 따라 획득하여, 상기 학생 총손실(L^S_{total})을 상기 학생 네트워크로 역전파할 수 있다.

[0033] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 학생 네트워크와 학습 시에 결합되는 교사 네트워크 및 학습부를 포함하는 영상 업스케일링 장치의 영상 스케일링 방법은 상기 학습부가 인코더와 디코더를 포함하는 오토 인코더로 구현되는 상기 교사 네트워크에 미리 획득된 학습 데이터 셋의 고해상도 영상을 입력하고, 학습된 방식에 따라 상기 고해상도 영상의 특징을 추출하여 합축 특징맵이 획득되면, 획득된 합축 특징맵으로부터 학습된 방식에 따라 고해상도 영상을 재구성하여 재구성 영상을 획득하고, 상기 학습 데이터 셋의 고해상도 영상과 대응하는 저해상도 영상, 상기 재구성 영상 및 상기 합축 특징맵을 기반으로 상기 교사 네트워크를 학습시키는 단계; 상기 교사 네트워크에 대한 학습이 완료되면, 학습된 교사 네트워크를 기반으로 지식 증류 기법에 따라 학생 네트워크에 상기 학습 데이터 셋의 저해상도 영상을 인가하여 상기 학생 네트워크를 학습시키는 단계; 및 상기 학생 네트워크에 대한 학습이 완료된 이후, 영상 업스케일링 동작 시에, 학생 네트워크가 업스케일링 대상이 되는 저해상도의 대상 영상을 인가받아 미리 학습된 방식에 따라 업스케일링하여 고해상도의 업스케일링 영상을 획득하는 단계를 포함한다.

발명의 효과

[0034] 따라서, 본 발명의 실시예에 따른 영상 업스케일링 장치 및 방법은 고해상도 영상을 사전 정보로 이용하여 미리 학습된 교사 네트워크로부터 지식 증류 기법으로 지식 정보가 전달되는 단순 구조의 학생 네트워크를 이용하여 사용자 단말에서도 효율적으로 저해상도 영상으로부터 고품질의 고해상도의 영상을 획득할 수 있도록 한다.

도면의 간단한 설명

[0035] 도 1은 본 발명의 일 실시예에 따른 영상 업스케일링 장치의 개략적 구조를 나타낸다.

도 2는 도 1의 영상 업스케일링 장치의 지식 증류 기법에 따른 동작을 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 영상 업스케일링 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0036] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0037] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0038] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"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", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0039] 도 1은 본 발명의 일 실시예에 따른 영상 업스케일링 장치의 개략적 구조를 나타내고, 도 2는 도 1의 영상 업스케일링 장치의 지식 증류 기법에 따른 동작을 설명하기 위한 도면이다.

[0040] 도 1 및 도 2를 참조하면, 본 실시예에 따른 영상 업스케일링 장치는 교사 네트워크(100), 학생 네트워크(200) 및 학습부(300)를 포함한다.

[0041] 교사 네트워크(100)는 인공 신경망으로 구현될 수 있으며, 특히 본 실시예에서 교사 네트워크는 입력 영상을 인가받아 특징을 추출하고, 추출된 특징을 이용하여 입력 영상을 재구성하는 오토 인코더(Auto-encoder)로 구현된

다.

[0042] 오토 인코더로 구현되는 교사 네트워크(100)는 고해상도(High Resolution: HR) 영상(Y)을 인가받아 특징을 추출하여 저차원의 함축 특징맵($\hat{\mathbf{X}}^T$)을 획득하는 인코더(110)와 인코더(110)에서 획득된 함축 특징맵($\hat{\mathbf{X}}^T$)으로부터 고해상도의 영상을 재구성하여 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득하는 디코더(120)를 포함할 수 있다.

[0043] 인코더(110)는 일 예로 다수의 컨볼루션 레이어와 활성화 함수인 ReLU 를 포함하는 인공 신경망으로 구현될 수 있으며, 여기서는 일 예로 4개의 컨볼루션 레이어와 PReLU(Parametric ReLU)를 포함하는 인공 신경망으로 구현되는 것으로 가정하였다. 그리고 다수의 컨볼루션 레이어 중 하나의 컨볼루션 레이어(예를 들면 2번째 컨볼루션 레이어)는 기지정된 스케일 팩터(s)에 따른 스트라이드(stride)를 갖고 나머지 컨볼루션 레이어는 1의 스트라이드를 갖도록 설정될 수 있다. 여기서 스케일 팩터(s)는 교사 네트워크(100)에 입력되는 고해상도 영상(Y)과 학생 네트워크(200)에 입력되는 저해상도 영상(X)의 해상도 비($Y = X^s$)에 대응하여 결정될 수 있다. 일 예로 고해상도 영상(Y)이 UHD(Ultra High Definition) 영상이고 저해상도 영상(X)이 FHD(Full High Definition) 영상인 경우, 스케일 팩터(s)는 2로 설정될 수 있다.

[0044] 인공 신경망으로 구현되는 인코더(110)는 학습부(300)에 의해 미리 학습될 수 있으며, 고해상도 영상(Y)이 인가되면 미리 학습된 방식에 따라 고해상도 영상(Y)의 특징을 추출하여 저해상도 영상(X)에 대응하는 크기의 함축 특징맵($\hat{\mathbf{X}}^T$)을 획득한다.

[0045] 고해상도 영상(Y)을 인가받아 함축 특징맵($\hat{\mathbf{X}}^T$)을 획득하는 인코더(110)의 동작 함수를 G^T 라고 하면, 함축 특징맵($\hat{\mathbf{X}}^T$)은 수학식 1로 표현될 수 있다.

수학식 1

$$\hat{\mathbf{X}}^T = G^T(\mathbf{Y})$$

[0047] 한편 디코더(120)는 다수의 컨볼루션 레이어와 디콘볼루션 레이어를 포함하여 구성될 수 있다. 디코더(120)는 FSRCNN과 유사하게 인가되는 함축 특징맵($\hat{\mathbf{X}}^T$)에 대해 각각 특징 추출, 축소, 맵핑 및 확장 기능을 수행하는 다수의 컨볼루션 레이어와 다수의 컨볼루션 레이어에서 최종적으로 출력되는 특징맵을 고해상도의 영상으로 재구성하는 디콘볼루션 레이어를 포함할 수 있다.

[0048] 그리고 디코더(120) 또한 인코더(110)와 마찬가지로 인공 신경망으로 구현되므로, 학습부(300)에 의해 미리 학습되어, 인코더(110)로부터 함축 특징맵($\hat{\mathbf{X}}^T$)이 인가되면, 미리 학습된 방식에 따라 함축 특징맵($\hat{\mathbf{X}}^T$)으로부터 고해상도 영상(Y)을 재구성한 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득한다.

[0049] 함축 특징맵($\hat{\mathbf{X}}^T$)을 인가받아 고해상도의 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득하는 디코더(120)의 동작 함수를 F^T 라고 하면, 재구성 영상($\hat{\mathbf{Y}}^T$)은 수학식 2로 표현될 수 있다.

수학식 2

$$\hat{\mathbf{Y}}^T = F^T(\hat{\mathbf{X}}^T)$$

[0051] 교사 네트워크(100)의 인코더(110)와 디코더(120) 각각은 학습이 수행됨에 따라 다수의 컨볼루션 레이어의 가중치가 업데이트된다.

[0052] 한편 학생 네트워크(200)는 교사 네트워크(100)의 디코더(120)와 동일하게 구성된다. 즉 학생 네트워크(200)는 교사 네트워크(100)의 디코더(120)와 마찬가지로 FSRCNN과 유사한 구조로 다수의 컨볼루션 레이어와 디콘볼루션

레이어를 포함하여 구성될 수 있다.

[0053] 그리고 학생 네트워크(200)는 교사 네트워크(100)의 학습이 완료된 이후, 학습 완료된 교사 네트워크(100)의 지식 정보를 기반으로 지식 증류(knowledge distillation) 기법으로 학습될 수 있다.

[0054] 미리 학습된 인공 신경망을 이용하여 다른 인공 신경망을 학습시키는 기법으로는 지식 증류 기법과 전이 학습(transfer learning) 기법이 알려져 있다. 이중 전이 학습 기법의 경우, 학습된 인공 신경망, 즉 교사 네트워크(100)의 디코더(120)의 가중치를 학생 네트워크(200)에 그대로 적용하여 추가 학습을 수행하는 방식이나, 교사 네트워크(100)와 학생 네트워크(200)가 완전하게 동일한 구조를 갖고 동일한 대상 데이터에 대한 처리만을 수행할 수 있다.

[0055] 그에 반해 지식 증류 기법은 학생 네트워크(200)가 미리 학습된 교사 네트워크(100)의 연산 과정에서 추출되는 중간 특징맵 또는 최종 결과물과 동일한 중간 특징맵 또는 최종 결과물을 도출할 수 있도록 학생 네트워크(200)를 별도로 학습시키는 기법이다. 그리고 지식 증류 기법에서는 사 네트워크(100)의 중간 특징맵(또는 최종 결과물)과 학생 네트워크(200)의 중간 특징맵(또는 최종 결과물) 사이의 차이를 나타내는 증류 손실을 정의하여 학생 네트워크(200)가 교사 네트워크(200)에 대응하여 학습될 수 있도록 한다. 따라서 학생 네트워크(200)는 교사 네트워크(100)와 구조가 일치하지 않을 수 있다. 이에 본 실시예에서 학생 네트워크(200)는 교사 네트워크(100)의 디코더(120)와 일치하는 구조를 갖는다.

[0056] 지식 증류 기법에 따른 학습에 대한 상세한 설명은 후술하도록 한다.

[0057] 한편 지식 증류 기법에 따라 학습 완료된 학생 네트워크(200)는 이후 실제 이용시에 업스케일링 대상이 되는 저해상도 영상(X)을 인가받고, 학습된 방식에 따라 인가된 저해상도 영상(X)을 업스케일링하여 고해상도의 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득한다.

[0058] 저해상도 영상(X)을 인가받아 고해상도의 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득하는 학생 네트워크(200)의 동작 함수를 F^S 라고 하면, 업스케일링 영상($\hat{\mathbf{Y}}^S$)은 수학적 식 3으로 표현될 수 있다.

수학적 식 3

[0059]
$$\hat{\mathbf{Y}}^S = F^S(\mathbf{X})$$

[0060] 여기서는 학생 네트워크(200)는 저해상도 영상(X)을 인가받아 고해상도의 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득하도록 구성되는 반면, 교사 네트워크(100)는 고해상도 영상(Y)을 학습을 위한 사전 정보(Privileged Information)로서 인가받아 특징을 추출하여 함축 특징맵($\hat{\mathbf{X}}^T$)을 획득하고, 함축 특징맵($\hat{\mathbf{X}}^T$)으로부터 다시 고해상도의 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득하도록 구성된다. 이는 상기한 바와 같이, 본 실시예에서는 학생 네트워크(200)가 지식 증류 기법으로 학습되므로, 교사 네트워크(100)와 상이하게 구성될 수 있기 때문이다.

[0061] 교사 네트워크(100)가 인코더(110)를 더 구비하여 사전 정보인 고해상도 영상(Y)에서 함축 특징맵($\hat{\mathbf{X}}^T$)을 획득하여 디코더(120)로 전달하는 경우, 디코더(120)는 고해상도 영상(Y)의 특징이 집약된 함축 특징맵($\hat{\mathbf{X}}^T$)으로부터 고품질의 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득하도록 학습될 수 있다. 그리고 디코더(120)가 고품질의 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득하도록 학습되면, 학습되는 학생 네트워크(200)는 지식 증류 기법에 따라 저해상도 영상(X)으로부터 고품질의 고해상도의 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득하도록 학습될 수 있다.

[0062] 학습부(300)는 교사 네트워크(100)를 학습시키기 위한 교사 학습부(310)와 학생 네트워크(200)를 학습시키기 위한 학생 학습부(320)를 포함할 수 있다.

[0063] 교사 학습부(310)는 교사 네트워크(100)의 인코더(110)와 디코더(120)의 학습 과정에서 모방 손실(L_{im}^T)과 복

원 손실($L_{\text{recon}}^{\mathcal{T}}$)을 계산하고, 계산된 모방 손실($L_{\text{im}}^{\mathcal{T}}$)과 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)에 따라 획득되는 교사 총손실($L_{\text{total}}^{\mathcal{T}}$)을 계산하여 교사 네트워크(100)로 역전파함으로써, 교사 네트워크(100)를 학습시킨다.

[0064] 여기서 모방 손실($L_{\text{im}}^{\mathcal{T}}$)은 학습 시에 교사 네트워크(100)의 인코더(110)에 고해상도 영상(Y)이 인가되어 획득된 함축 특징맵($\hat{\mathbf{X}}^{\mathcal{T}}$)과 고해상도 영상(Y)에 대응하는 저해상도 영상(X) 사이의 차로 수학식 4와 같이 계산될 수 있다.

수학식 4

$$L_{\text{im}}^{\mathcal{T}} = \frac{1}{H'W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} |X_{ij} - \hat{X}_{ij}^{\mathcal{T}}|$$

[0066] (여기서 i, j는 픽셀 위치 좌표를 나타내고, H' 과 W' 은 각각 저해상도 영상(X)의 높이와 폭을 나타낸다.)

[0067] 즉 모방 손실($L_{\text{im}}^{\mathcal{T}}$)은 교사 네트워크(100)의 인코더(110)가 고해상도 영상(Y)으로부터 얼마나 저해상도 영상(X)과 유사한 함축 특징맵($\hat{\mathbf{X}}^{\mathcal{T}}$)을 획득할 수 있는지를 나타낸다. 여기서 교사 네트워크(100)의 학습 시에 이용되는 저해상도 영상(X)은 사전 정보인 고해상도 영상(Y)에서 해상도만을 기존의 다양한 다운샘플링 기법으로 저하시킨 영상이다. 영상에 대한 다운샘플링 기법은 업샘플링 기법에 비해 간단한 연산으로 매우 용이하게 획득될 수 있으며, 다양한 기법이 공지되어 있으므로 여기서는 상세하게 설명하지 않는다. 즉 교사 네트워크(100)의 학습 시에는 고해상도 영상(Y)과 고해상도 영상(Y)을 다운 샘플링한 저해상도 영상(X)이 학습 데이터셋으로 이용될 수 있다.

[0068] 한편, 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)은 교사 네트워크(100)가 고해상도 영상(Y)을 함축 특징맵($\hat{\mathbf{X}}^{\mathcal{T}}$)으로 함축한 후, 재구성 영상($\hat{\mathbf{Y}}^{\mathcal{T}}$)으로 재구성시에 얼마나 고해상도 영상(Y)과 유사하게 복원할 수 있는지를 나타내는 손실로서, 고해상도 영상(Y)과 재구성 영상($\hat{\mathbf{Y}}^{\mathcal{T}}$) 사이의 차로 수학식 5와 같이 계산될 수 있다.

수학식 5

$$L_{\text{recon}}^{\mathcal{T}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^{\mathcal{T}}|$$

[0070] (여기서 i, j는 픽셀 위치 좌표를 나타내고, H 과 W 은 각각 고해상도 영상(Y)의 높이와 폭을 나타낸다.)

[0071] 이에 교사 학습부(310)는 모방 손실($L_{\text{im}}^{\mathcal{T}}$)과 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)의 합으로 수학식 6에 따라 교사 총손실($L_{\text{total}}^{\mathcal{T}}$)을 획득할 수 있다.

수학식 6

$$L_{\text{total}}^{\mathcal{T}} = L_{\text{recon}}^{\mathcal{T}} + \lambda^{\mathcal{T}} L_{\text{im}}^{\mathcal{T}}$$

[0073] (여기서 $\lambda^{\mathcal{T}}$ 는 모방 손실($L_{\text{im}}^{\mathcal{T}}$)과 복원 손실($L_{\text{recon}}^{\mathcal{T}}$)의 비중을 조절하기 위한 파라미터이다.)

[0074] 교사 학습부(310)는 획득된 교사 총손실(L_{total}^T)을 교사 네트워크로 역전파함으로써, 인코더(110)와 디코더(120)의 가중치, 즉 지식 정보를 업데이트함으로써 학습을 수행할 수 있다. 이때, 교사 학습부(310)는 기지정된 횟수 또는 교사 총손실(L_{total}^T)이 기지정된 기준 교사 총손실 이하가 되도록 반복 학습을 수행하여, 인코더(110)와 디코더(120)의 지식 정보를 요구되는 수준으로 업데이트 할 수 있다.

[0075] 한편, 학생 학습부(320)는 교사 학습부(310)의 교사 네트워크(100) 학습 완료 여부를 확인하고, 교사 네트워크(100)의 학습이 완료되면, 학습된 디코더(120)를 기반으로 학생 네트워크(200)로 지식 증류 기법에 따라 학습시킨다. 이때 학생 학습부(320)는 학생 네트워크(200)가 함축 특징맵($\hat{\mathbf{X}}^T$)이 아닌 저해상도 영상(X)으로부터 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득할 수 있도록 증류 학습 기법에 따라 학습을 수행한다.

[0076] 학생 학습부(320)는 우선 학습 데이터 셋의 고해상도 영상(Y)과 업스케일링 영상($\hat{\mathbf{Y}}^S$) 사이의 차에 따라 수학적 식 7에 따라 학생 네트워크(200)의 복원 손실(L_{recon}^S)을 수학적 식 7과 같이 계산할 수 있다.

수학적 식 7

$$L_{recon}^S = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W |Y_{ij} - \hat{Y}_{ij}^S|$$

[0078] 여기서 복원 손실(L_{recon}^S)은 학생 네트워크(200)가 저해상도 영상(X)에 대응하는 고해상도 영상(Y)을 얼마나 정확하게 복원하였는지 여부를 나타낸다.

[0079] 이하에서는 학생 학습부(320)에서 획득되는 복원 손실(L_{recon}^S)을 교사 학습부(320)에서 획득되는 복원 손실(L_{recon}^T)과 구분하기 위해 교사 복원 손실(L_{recon}^T)과 학생 복원 손실(L_{recon}^S)이라 한다.

[0080] 또한 학생 학습부(320)는 학생 네트워크(200)가 지식 증류 기법에 따라 학습된 교사 네트워크(100)의 지식을 습득할 수 있도록 한다. 본 실시예에서 학생 학습부(320)는 동일한 구조를 갖는 디코더(120)와 학생 네트워크(200)의 다수의 컨볼루션 레이어 중 서로 대응하는 동일 위치의 하나의 컨볼루션 레이어에서 출력되는 중간 특징맵(f^T , f^S)을 획득하고, 획득된 중간 특징맵(f^T , f^S) 사이의 상호 정보($I(f^T; f^S)$)가 최대가 되도록 하는 증류 손실(distillation loss)($L_{distill}^S$)을 정의하고, 정의된 증류 손실($L_{distill}^S$)을 이용하여 학생 네트워크(200)가 지식 증류 기법에 따라 학습되도록 한다.

[0081] 여기서 중간 특징맵(f^T , f^S) 사이의 상호 정보($I(f^T; f^S)$)는 수학적 식 8로 정의될 수 있다.

수학적 식 8

$$I(f^T; f^S) = H(f^T) - H(f^T | f^S)$$

[0083] (여기서 $H(f^T)$ 는 주변 엔트로피(marginal entropy)를 나타내고, $H(f^T | f^S)$ 는 조건부 엔트로피(conditional entropy)를 나타낸다.)

[0084] 수학적 식 8로부터 상호 정보($I(f^T; f^S)$)가 최대화되기 위해서는 조건부 엔트로피($H(f^T | f^S)$)를 최소화해야 한다. 그러나 조건부 확률($p(f^T | f^S)$)을 최소화하는 것은 용이하지 않다. 이에 학생 학습부(320)는 수학적 식 8의 변동 정보 최대화(variational information maximization) 기법에 따라 가우시안 분포(Gaussian distribution)나 라플라스 분포(Laplace distribution)와 같은 중간 특징맵(f^T , f^S) 사이의 분포 관계에 대한 파라메트릭 모델

(parametric model)($q(f^T; f^S)$)을 추정하여 조건부 분포($p(f^T|f^S)$)를 근사화함으로써, 상호 정보($I(f^T; f^S)$)의 하한을 계산할 수 있으며, 상호 정보($I(f^T; f^S)$)의 하한이 최대화되도록 할 수 있다.

[0085] 여기서의 일 예로 학생 학습부(320)가 다변량 라플라스 분포(multivariate Laplace distribution)를 기반으로 하는 중간 특징맵(f^T, f^S) 사이의 파라메트릭 모델($q(f^T; f^S)$)을 이용하여 중간 특징맵(f^T, f^S) 사이의 픽셀 위치(location)에 대한 파라메트릭인 위치맵($\mu \in \mathbb{R}^{C \times H' \times W'}$, 여기서 R은 실수를 나타내고, C는 중간 특징맵(f^T, f^S)의 채널 수)과 스케일(scale)에 대한 파라메트릭인 스케일맵($b \in \mathbb{R}^{C \times H' \times W'}$)을 추정하는 것으로 가정한다.

[0086] 따라서 본 실시예에 따른 학생 학습부(320)는 학생 네트워크의 중간 특징맵(f^S)로부터 파라메트릭 모델($q(f^T; f^S)$)에 기반하여 위치맵(μ)과 스케일맵(b)을 추정하고, 추정된 위치맵(μ)과 스케일맵(b) 및 교사 네트워크의 중간 특징맵(f^T)을 이용하여 중간 특징맵(f^T, f^S) 사이의 상호 정보($I(f^T; f^S)$)의 하한이 최대가 되도록 하는 증류 손실(L^S_{distill})을 수학식 9로 정의한다.

수학식 9

$$L^S_{\text{distill}} = \frac{1}{CH'W'} \sum_{i=1}^C \sum_{j=1}^{H'} \sum_{k=1}^{W'} \log b_{ijk} + \frac{|f^{\mathcal{T}}_{ijk} - \mu_{ijk}|}{b_{ijk}}$$

[0088] 수학식 9에 따르면, 증류 손실(L^S_{distill})이 감소될수록 교사 네트워크의 중간 특징맵(f^T)과 위치맵(μ) 사이의 거리가 최소화되며, 스케일맵(b)은 증류 범위를 제어한다.

[0089] 그리고 학생 학습부(320)는 학생 복원 손실(L^S_{recon})과 증류 손실(L^S_{distill})의 합으로 수학식 10에 따라 학생 총손실(L^S_{total})을 획득할 수 있다.

수학식 10

$$L^S_{\text{total}} = L^S_{\text{recon}} + \lambda^S L^S_{\text{distill}}$$

[0091] (여기서 λ^S 는 학생 복원 손실(L^S_{recon})과 증류 손실(L^S_{distill})의 비중을 조절하기 위한 파라미터이다.)

[0092] 학생 학습부(320)는 획득된 학생 총손실(L^S_{total})을 학생 네트워크(200)로 역전파하여 학생 네트워크(200)에 대한 학습을 수행할 수 있다. 학생 학습부(320) 또한 기지정된 횟수 또는 학생 총손실(L^S_{total})이 기지정된 기준 학생 총손실 이하가 되도록 반복 학습을 수행할 수 있다.

[0093] 도 1에 도시된 영상 업스케일링 장치에서 교사 네트워크(100)와 학습부(300)는 학생 네트워크(200)를 학습시키기 위한 구성으로 고성능의 서버 등에서 학생 네트워크(200)와 함께 학습될 수 있으나, 학생 네트워크(200)의 학습이 완료되면 제외될 수 있다. 즉 영상 업스케일링 장치의 실제 이용시에는 학생 네트워크만이 이용된다.

[0094] 상기한 바와 같이, 사전 정보를 이용한 증류 기법 기반 학습이 완료된 학생 네트워크(200)는 저해상도 영상으로부터 기존의 FSRCNN이나 교사 네트워크와 동일한 구조를 갖는 학생 네트워크가 증류 기법으로 학습되는 기존의 방식에 비해 매우 고품질의 고해상도 영상을 획득할 수 있다.

[0095] 상기에서는 오토 인코더로 구현되는 교사 네트워크(100)와 교사 네트워크(100)의 디코더(120)와 동일한 구조를

갖는 학생 네트워크(200)가 다수의 컨볼루션 레이어를 포함하는 것으로 설명하였으나, 이는 교사 네트워크(100)와 학생 네트워크(200)가 컨볼루션 연산에 기반하여 동작하는 경우에 대한 예시로서, 경우에 따라서는 다른 연산을 수행하는 레이어가 이용될 수도 있다. 즉 컨볼루션 레이어는 교사 네트워크(100)와 학생 네트워크(200)의 구성에 따라 다양한 연산 레이어로 구성될 수 있다.

[0096] 도 3은 본 발명의 일 실시예에 따른 영상 업스케일링 방법을 나타낸다.

[0097] 도 1 및 도 2를 참조하여, 도 3의 영상 업스케일링 방법을 설명하면, 우선 학습부(300)가 인코더(110)와 디코더(120)를 포함하는 오토 인코더로 구현되는 교사 네트워크(100)를 학습시킨다.

[0098] 교사 네트워크를 학습시키는 단계에서는 먼저 미리 준비된 학습 데이터 셋의 고해상도 영상(Y)을 사전 정보로서 입력 받는다(S11). 여기서 학습 데이터 셋은 고해상도 영상(Y)과 이를 다운스케일링하여 미리 획득된 저해상도 영상(X)의 집합이다. 그리고 교사 네트워크(100)의 인코더(110)는 현재까지 학습된 방식에 따라 고해상도 영상(Y)의 특징을 추출하여, 함축 특징맵($\hat{\mathbf{X}}^T$)을 획득한다(S12). 함축 특징맵($\hat{\mathbf{X}}^T$)이 획득되면, 교사 네트워크(100)의 디코더(120)가 현재까지 학습된 방식에 따라 획득된 함축 특징맵($\hat{\mathbf{X}}^T$)으로부터 고해상도의 영상을 재구성하여 재구성 영상($\hat{\mathbf{Y}}^T$)을 획득한다(S13).

[0099] 이에 학습부(300)는 획득된 함축 특징맵($\hat{\mathbf{X}}^T$)과 저해상도 영상(X)을 이용하여, 수학적 4에 따라 모방 손실(L_{im}^T)을 계산하고, 재구성 영상($\hat{\mathbf{Y}}^T$)과 고해상도 영상(Y)을 이용하여 교사 복원 손실(L_{recon}^T)을 수학적 5에 따라 계산한다(S14). 그리고 계산된 모방 손실(L_{im}^T)과 교사 복원 손실(L_{recon}^T)을 합하여, 수학적 6과 같이 교사 총손실(L_{total}^T)을 획득한다(S15).

[0100] 이후, 학습부(300)는 교사 네트워크(100)에 대한 학습이 완료되었는지 판별한다(S16). 여기서 학습부(300)는 기지정된 횟수만큼 학습이 반복되었거나, 교사 총손실(L_{total}^T)이 기지정된 기준 교사 총손실 이하가 되면, 교사 네트워크(100)에 대한 학습이 완료된 것으로 판별할 수 있다.

[0101] 만일 교사 네트워크(100)에 대한 학습이 완료되지 않았으면, 학습부(300)는 교사 총손실(L_{total}^T)을 교사 네트워크(100)로 역전파하여 교사 네트워크(100)의 지식 정보인 가중치를 업데이트 한다(S17).

[0102] 그러나 교사 네트워크(100)에 대한 학습이 완료된 것으로 판별되면, 학습부(300)는 지식 증류 기법을 기반으로 학생 네트워크(200)에 대한 학습을 수행한다.

[0103] 학생 네트워크 학습 시에 학습부(300)는 학생 네트워크(200)에 학습 데이터 셋의 저해상도 영상(X)을 입력시킨다(S21). 이에 학생 네트워크(200)는 인가된 저해상도 영상(X)을 학습되는 방식에 따라 업스케일링하여 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득한다(S22). 학습부(300)는 업스케일링 영상($\hat{\mathbf{Y}}^S$)을 획득하는 과정에서 학생 네트워크(200)를 구성하는 다수의 레이어 중 기지정된 레이어에서 출력되는 중간 특징맵(f^S)을 인가받고, 인가된 파라메트릭 모델($q(f^T; f^S)$)에 기반하여 중간 특징맵(f^S)에 대한 위치맵(μ)과 스케일맵(b)을 추정한다(S23).

[0104] 위치맵(μ)과 스케일맵(b)이 추정되면, 학습부(300)는 증류 손실(L_{distill}^S)을 수학적 9에 따라 계산하고, 학생 복원 손실(L_{recon}^S)을 수학적 7에 따라 계산한다(S24). 그리고 계산된 증류 손실(L_{distill}^S)과 학생 복원 손실(L_{recon}^S)로부터 학생 총손실(L_{total}^S)을 수학적 10에 따라 계산하여 획득한다(S25).

[0105] 학습부(300)는 학생 네트워크(200)에 대한 학습이 완료되었는지 판별한다(S26). 학습부(300)는 학생 네트워크(200)에 대한 학습이 기지정된 횟수만큼 학습이 반복되었거나, 학생 총손실(L_{total}^S)이 기지정된 기준 학생 총손실 이하가 되면, 학생 총손실(L_{total}^S)에 대한 학습이 완료된 것으로 판별할 수 있다.

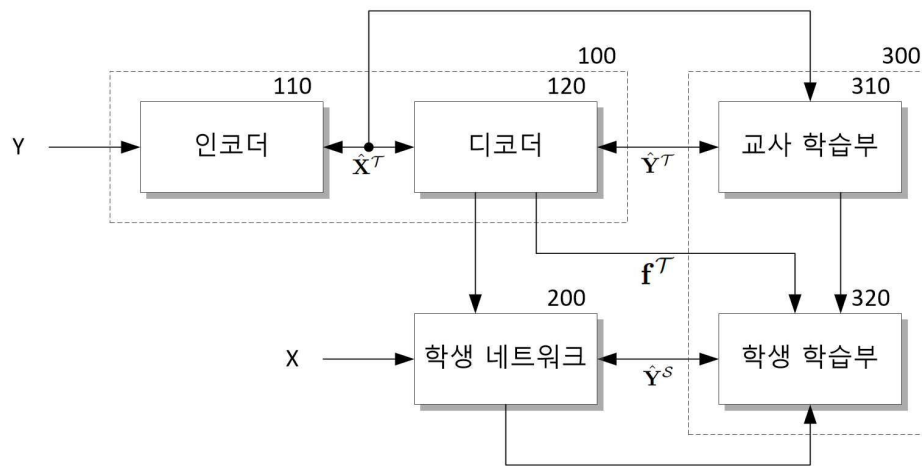
- [0106] 만일 학생 네트워크(200)에 대한 학습이 완료되지 않았으면, 학습부(300)는 학생 총손실(L_{total}^S)을 학생 네트워크(200)로 역전파하여 학생 네트워크(200)의 가중치를 업데이트 한다(S27). 이때, 학생 네트워크(200)가 학생 복원 손실(L_{recon}^S)뿐만 아니라, 증류 손실($L_{distill}^S$)로부터 학생 총손실(L_{total}^S)을 획득하므로, 학생 네트워크(200)는 미리 학습된 교사 네트워크(100)로부터 지식 증류 기법에 따라 지식을 전달받아 학습되는 것으로 볼 수 있다.
- [0107] 한편 학생 네트워크(200)에 대한 학습이 완료되었으면, 학생 네트워크(200)를 업스케일링 작업을 수행해야 하는 장치에 실장 시킨다(S31). 그리고 업스케일링 대상이 되는 저해상도의 대상 영상을 입력받아 미리 학습된 방식에 따라 업스케일링하여 고해상도의 업스케일링 영상(\hat{Y}^S)을 출력한다(S32).
- [0108] 결과적으로 본 실시예에 따른 영상 업스케일링 장치 및 방법은 교사 네트워크(100)를 저해상도 영상(X)에서 곧바로 고해상도의 재구성 영상(\hat{Y}^T)을 획득하도록 학습시키지 않고, 고해상도 영상(Y)에서 함축 특징맵(\hat{X}^T)을 획득한 후, 함축 특징맵(\hat{X}^T)에서 재구성 영상(\hat{Y}^T)을 획득하도록 학습시킨 후, 학습된 교사 네트워크(100)를 기반으로 지식 증류 기법으로 학생 네트워크(200)를 학습 시킴으로써, 학생 네트워크(200)가 간단한 구조로도 매우 높은 수준의 업스케일링 영상(\hat{Y}^S)을 재구성하도록 학습시킬 수 있다. 따라서 하드웨어 성능의 제약이 많은 사용자 단말에서도 학생 네트워크(200)만을 구비하여 저해상도 영상으로부터 고품질의 고해상도 영상이 재구성되도록 할 수 있다.
- [0109] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.
- [0110] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.
- [0111] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

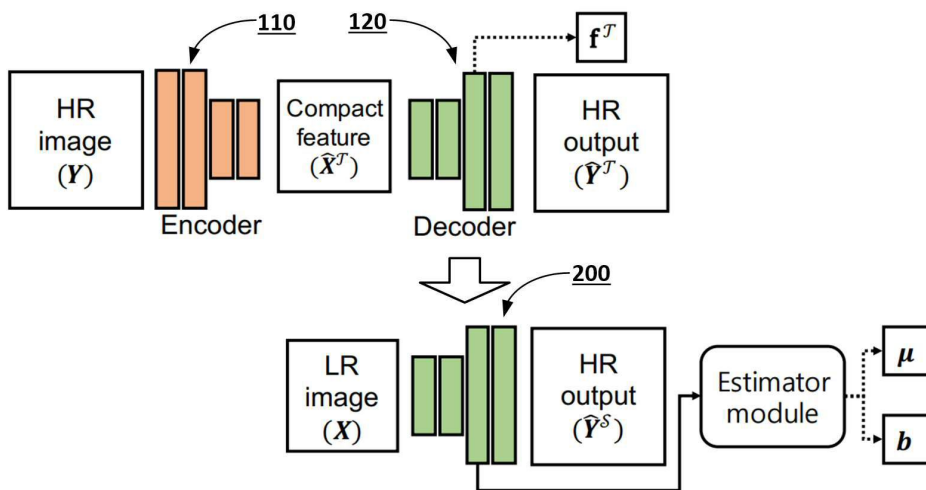
- [0112] 100: 교사 네트워크 110: 인코더
120: 디코더 200: 학생 네트워크
300: 학습부 310: 교사 학습부
320: 학생 학습부

도면

도면1



도면2



도면3

