(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0052314
(43) 공개일자 2025년04월18일

(51) 국제특허분류(Int. Cl.)
H04N 19/30 (2014.01) *G06N 3/045* (2023.01)
G06N 3/0464 (2023.01) *G06N 3/084* (2023.01)
H04N 19/124 (2014.01) *H04N 19/29* (2014.01)
H04N 19/91 (2014.01)
(52) CPC특허분류
H04N 19/30 (2015.01)
G06N 3/045 (2023.01)
(21) 출원번호 10-2024-0138967
(22) 출원일자 2024년10월11일
심사청구일자 없음
(30) 우선권주장
1020230135463 2023년10월11일 대한민국(KR)
(뒷면에 계속)

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
이주영
대전광역시 유성구 가정로 218
김연희
대전광역시 유성구 가정로 218
(뒷면에 계속)
(74) 대리인
최윤서

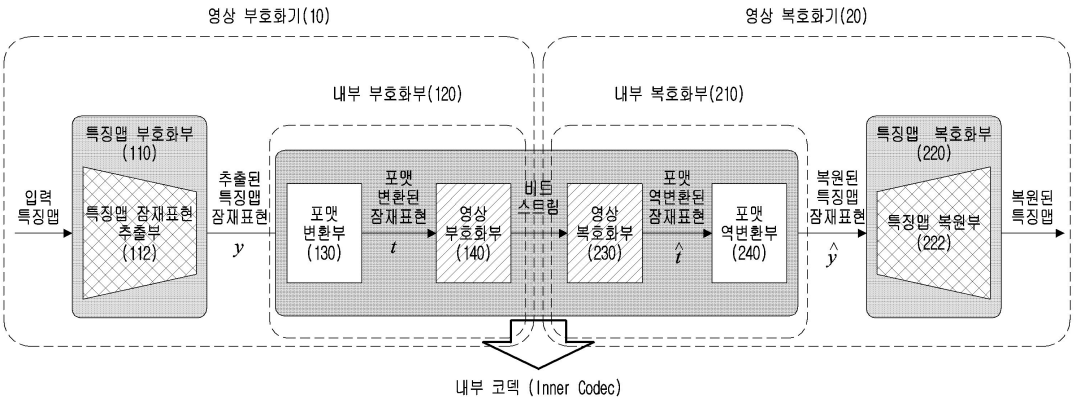
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **특징맵 부호화/복호화를 위한 장치 및 이를 위한 방법**

(57) 요 약

본 개시에 따른 특징맵 복호화를 위한 장치는, 비트스트림으로부터, 영상을 복호화하는, 영상 복호화부; 복호화된 영상의 포맷을 변환하여, 특징맵 잠재 표현을 복원하는 포맷 역변환부; 및 복호화된 특징맵 잠재 표현으로부터, 다계층 특징맵을 복원하는 특징맵 복원부를 포함할 수 있다. 이때, 상기 특징맵 복원부는, 학습된 신경망 파라미터를 기초로, 상기 다계층 특징맵을 복원할 수 있다.

대 표 도



(52) CPC특허분류
G06N 3/0464 (2023.01)
G06N 3/084 (2023.01)
H04N 19/124 (2015.01)
H04N 19/29 (2015.01)
H04N 19/91 (2015.01)

(72) 발명자
정세윤
대전광역시 유성구 가정로 218
최진수
대전광역시 유성구 가정로 218
강정원
대전광역시 유성구 가정로 218
김휘용
경기도 용인시 기흥구 덕영대로 1732, 경희대학교
정혜원
경기도 용인시 기흥구 덕영대로 1732, 경희대학교
장승환
경기도 용인시 기흥구 덕영대로 1732, 경희대학교
김영웅
경기도 용인시 기흥구 덕영대로 1732, 경희대학교
유장현
경기도 용인시 기흥구 덕영대로 1732, 경희대학교

(30) 우선권주장
1020240021365 2024년02월14일 대한민국(KR)
1020240053699 2024년04월22일 대한민국(KR)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711193547
과제번호 2020-0-00011
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 ETRI연구개발지원사업
연구과제명 (전문연구실) 기계를 위한 영상 부호화 기술
과제수행기관명 ETRI
연구기간 2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

비트스트림으로부터, 영상을 복호화하는, 영상 복호화부;

복호화된 영상의 포맷을 변환하여, 특징맵 잠재 표현을 복원하는 포맷 역변환부; 및

복호화된 특징맵 잠재 표현으로부터, 다계층 특징맵을 복원하는 특징맵 복원부를 포함하되,

상기 특징맵 복원부는, 학습된 신경망 파라미터를 기초로, 상기 다계층 특징맵을 복원하는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 2

제1 항에 있어서,

상기 특징맵 복원부는,

상기 영상 복호화부에서 이용 가능한 압축 파라미터들을 기반으로 학습되는 압축 파라미터 유관 특징맵 복원부 또는 상기 영상 복호화부에서 이용 가능한 압축 파라미터들을 고려함이 없이 학습되는 압축 파라미터 무관 특징맵 복원부 중 적어도 하나를 포함하는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 3

제2 항에 있어서,

상기 압축 파라미터는, 양자화 파라미터인 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 4

제2 항에 있어서,

상기 압축 파라미터 유관 특징맵 복원부는,

상기 압축 파라미터들을 기반으로 학습되는 압축 파라미터 적응부; 및

상기 압축 파라미터 적응부에 입력되는 채널들의 개수를 조절하는 채널 적응부를 포함하는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 5

제4 항에 있어서,

상기 채널 적응부는, 상기 복호화된 특징맵 잠재 표현 및 상기 압축 파라미터들을 채널 결합하여 결합 영상을, 상기 복호화된 특징맵 잠재 표현의 채널 수에 맞춰 변환하는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 6

제1 항에 있어서,

상기 특징맵 복원부는, 오류 역전파로 학습 가능한 엔트로피 부호화부 및 엔트로피 복호화부를 포함하는 제1 내부 코덱의 압축 잡음을 기반으로 1차적으로 학습되는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 7

제6 항에 있어서,

1차적으로 학습된 신경망 파라미터는, 오류 역전파로 학습 불가능한 영상 부호화부 및 영상 복호화부를 포함하는 제2 내부 코덱의 압축 작음을 기반으로 미세 조정되는 것을 특징으로 하는, 특징맵 복호화를 위한 장치

청구항 8

제4 항에 있어서,

상기 포맷 역변환부는, 상기 복호화된 영상에 대해 역양자화를 수행하기 위한 역양자화부; 및

역양자화된 영상에 대해 채널 재배열을 수행하기 위한 채널 재배열부를 포함하는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 9

제7 항에 있어서,

상기 역양자화는, 특징값들 중 최대값 및 최소값에 기반하여 수행되고,

상기 최대값 및 상기 최소값을 나타내는 정보가 상기 비트스트림을 통해 명시적으로 복호화되는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 10

제8 항에 있어서,

상기 채널 재배열은, 공간적, 시간적 또는 시공간적으로 배열된 채널들을, 원래 형태로 복원하는 것을 특징으로 하는, 특징맵 복호화를 위한 장치.

청구항 11

입력되는 다계층 특징맵으로부터, 특징맵 잠재 표현을 추출하는, 특징맵 잠재 표현 추출부;

상기 특징맵 잠재 표현의 포맷을 변환하는, 포맷 변환부; 및

포맷 변환된 영상을 부호화하여 비트스트림을 생성하는, 영상 부호화부를 포함하되,

상기 특징맵 잠재 표현 추출부는, 학습된 신경망 파라미터를 기초로, 상기 특징맵 잠재 표현을 추출하는 것을 특징으로 하는, 특징맵 부호화를 위한 장치.

청구항 12

제11 항에 있어서,

상기 특징맵 잠재 표현에 대해 채널 변환을 수행하기 위한 채널 배열부; 및

채널 변환된 영상에 대해 양자화를 수행하기 위한 양자화부를 포함하는 것을 특징으로 하는, 특징맵 부호화를 위한 장치.

청구항 13

제12 항에 있어서,

상기 양자화는, 특징값들 중 최대값 및 최소값에 기반하여 수행되고,

상기 최대값 및 상기 최소값을 나타내는 정보가 비트스트림을 통해 명시적으로 시그날링되는 것을 특징으로 하는, 특징맵 부호화를 위한 장치.

청구항 14

제8 항에 있어서,

상기 채널 배열은, 상기 특징맵 잠재 표현의 채널들을, 공간적, 시간적 또는 시공간적으로 배열하는 것을 특징으로 하는, 특징맵 부호화를 위한 장치.

청구항 15

비트스트림으로부터, 영상을 복호화하는 단계;

복호화된 영상의 포맷을 변환하여, 특징맵 잠재 표현을 복원하는 단계; 및

복호화된 특징맵 잠재 표현으로부터, 다계층 특징맵을 복원하는 단계를 포함하되,

상기 특징맵을 복원하는 단계는, 학습된 신경망 파라미터를 기초로 수행되는 것을 특징으로 하는, 특징맵 복호화 방법.

발명의 설명

기 술 분 야

[0001] 본 개시는, 특징맵 부호화/복호화를 위한 장치 및 이를 위한 방법에 관한 것이다.

배 경 기 술

[0002] 전통적인 영상 압축 기술은, 압축된 영상을 복원하였을 때, 인간의 시각을 기준으로, 복원된 영상이 원본과 최대한 유사해지도록 발전하여 왔다. 즉, 영상 압축 기술은, 비트량을 최소화함과 동시에 복원 영상의 화질을 최대화하는 방향으로 발전되었다.

[0003] 일 예로, 인코더는, 영상을 입력으로 받아, 입력된 영상에 대한 변환 및 엔트로피 부호화 과정을 통해 비트스트림을 생성하고, 디코더는, 비트스트림을 입력 받아, 다시 원본과 유사한 영상으로 복원한다.

[0004] 원본 영상과 복원 영상간의 유사도를 측정하기 위해, 객관적 화질 평가 척도 또는 주관적 화질 평가 척도를 사용할 수 있다. 여기서, 객관적 화질 평가 척도로서, 원본 영상과 복원 영상 사이의 픽셀 값들의 차이를 측정하는 평균오차제곱(MSE) 등이 주로 이용된다. 한편, 주관적 화질 평가 척도는, 원본 영상과 복원 영상 사이의 차이를 사람이 평가하는 것이다.

[0005] 한편, 기계 시각 작업 성능이 향상되면서, 사람 대신, 기계가 영상을 보고 소비하는 경우가 증가하고 있다. 일 예로, 스마트 시티, 자율주행 자동차, 공항 감시 카메라 등의 분야에서는, 인간 보다는, 기계를 중심으로, 영상을 소비하는 경우가 증가하고 있다.

[0006] 이에 따라, 최근에는, 인간을 중심으로하는 전통적인 영상 압축 이외에, 기계 시각(Machine Vision)을 중심으로 하는 영상 압축 기술에 대한 관심이 증가하고 있다.

발명의 내용

해결하려는 과제

[0007] 본 개시는, 다계층 특징맵으로부터 특징맵 잠재 표현을 추출하고, 특징맵 잠재 표현을 다계층 특징맵으로 복원하는 방법을 제공한다.

[0008] 본 개시는, 특징맵 잠재 표현의 추출 및 다계층 특징맵을 복원함에 있어서, 학습된 신경망 파라미터를 이용하는 방법을 제공한다.

[0009] 본 개시는, 압축 파라미터를 기반으로 신경망 파라미터를 학습하는 방법을 제공한다.

[0010] 본 개시는, 일반적인 코덱을 이용하여 부호화/복호화할 수 있도록, 특징맵 잠재 표현의 포맷을 변환/역변환하는 방법을 제공한다.

[0011] 본 개시에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0012] 본 개시에 따른 특징맵 복호화를 위한 장치는, 비트스트림으로부터, 영상을 복호화하는, 영상 복호화부; 복호화된 영상의 포맷을 변환하여, 특징맵 잠재 표현을 복원하는 포맷 역변환부; 및 복호화된 특징맵 잠재 표현으로부터, 다계층 특징맵을 복원하는 특징맵 복원부를 포함할 수 있다. 이때, 상기 특징맵 복원부는, 학습된 신경망

파라미터를 기초로, 상기 다계층 특징맵을 복원할 수 있다.

[0013] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 특징맵 복원부는, 상기 영상 복호화부에서 이용 가능한 압축 파라미터들을 기반으로 학습되는 압축 파라미터 유관 특징맵 복원부 또는 상기 영상 복호화부에서 이용 가능한 압축 파라미터들을 고려함이 없이 학습되는 압축 파라미터 무관 특징맵 복원부 중 적어도 하나를 포함할 수 있다.

[0014] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 압축 파라미터는, 양자화 파라미터일 수 있다.

[0015] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 압축 파라미터 유관 특징맵 복원부는, 상기 압축 파라미터들을 기반으로 학습되는 압축 파라미터 적응부; 및 상기 압축 파라미터 적응부에 입력되는 채널들의 개수를 조절하는 채널 적응부를 포함할 수 있다.

[0016] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 채널 적응부는, 상기 복호화된 특징맵 잠재 표현 및 상기 압축 파라미터들을 채널 결합하여 결합 영상을, 상기 복호화된 특징맵 잠재 표현의 채널 수에 맞춰 변환할 수 있다.

[0017] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 특징맵 복원부는, 오류 역전파로 학습 가능한 엔트로피 부호화부 및 엔트로피 복호화부를 포함하는 제1 내부 코덱의 압축 잡음을 기반으로 1차적으로 학습될 수 있다.

[0018] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 1차적으로 학습된 신경망 파라미터는, 오류 역전파로 학습 불가능한 영상 부호화부 및 영상 복호화부를 포함하는 제2 내부 코덱의 압축 작음을 기반으로 미세 조정될 수 있다.

[0019] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 포맷 역변환부는, 상기 복호화된 영상에 대해 역양자화를 수행하기 위한 역양자화부; 및 역양자화된 영상에 대해 채널 재배열을 수행하기 위한 채널 재배열부를 포함할 수 있다.

[0020] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 역양자화는, 특징값들 중 최대값 및 최소값에 기반하여 수행되고, 상기 최대값 및 상기 최소값을 나타내는 정보가 상기 비트스트림을 통해 명시적으로 복호화될 수 있다.

[0021] 본 개시에 따른 특징맵 복호화를 위한 장치에 있어서, 상기 채널 재배열은, 공간적, 시간적 또는 시공간적으로 배열된 채널들을, 원래 형태로 복원하는 것일 수 있다.

[0022] 본 개시에 따른 특징맵 부호화를 위한 장치는, 입력되는 다계층 특징맵으로부터, 특징맵 잠재 표현을 추출하는, 특징맵 잠재 표현 추출부; 상기 특징맵 잠재 표현의 포맷을 변환하는, 포맷 변환부; 및 포맷 변환된 영상을 부호화하여 비트스트림을 생성하는, 영상 부호화부를 포함할 수 있다. 이때, 상기 특징맵 잠재 표현 추출부는, 학습된 신경망 파라미터를 기초로, 상기 특징맵 잠재 표현을 추출할 수 있다.

[0023] 본 개시에 따른 특징맵 부호화를 위한 장치에 있어서, 상기 특징맵 잠재 표현에 대해 채널 변환을 수행하기 위한 채널 배열부; 및 채널 변환된 영상에 대해 양자화를 수행하기 위한 양자화부를 포함할 수 있다.

[0024] 본 개시에 따른 특징맵 부호화를 위한 장치에 있어서, 상기 양자화는, 특징값들 중 최대값 및 최소값에 기반하여 수행되고, 상기 최대값 및 상기 최소값을 나타내는 정보가 비트스트림을 통해 명시적으로 시그날링될 수 있다.

[0025] 본 개시에 따른 특징맵 부호화를 위한 장치에 있어서, 상기 채널 배열은, 상기 특징맵 잠재 표현의 채널들을, 공간적, 시간적 또는 시공간적으로 배열하는 것일 수 있다.

[0026] 본 개시에 따른 특징맵 복호화 방법은, 비트스트림으로부터, 영상을 복호화하는 단계; 복호화된 영상의 포맷을 변환하여, 특징맵 잠재 표현을 복원하는 단계; 및 복호화된 특징맵 잠재 표현으로부터, 다계층 특징맵을 복원하는 단계를 포함할 수 있다. 이때, 상기 특징맵을 복원하는 단계는, 학습된 신경망 파라미터를 기초로 수행될 수 있다.

[0027] 본 개시에 따르면, 특징맵 부호화/복호화 방법을 기록한 컴퓨터로 판독 가능한 기록 매체가 제공될 수 있다.

[0028] 본 개시에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하

게 이해될 수 있을 것이다.

발명의 효과

[0029] 본 개시에 의하면, 다계층 특징맵으로부터 특징맵 잠재 표현을 추출하고, 특징맵 잠재 표현을 다계층 특징맵으로 복원하는 방법을 제공할 수 있다.

[0030] 본 개시에 의하면, 특징맵 잠재 표현의 추출 및 다계층 특징맵을 복원함에 있어서, 학습된 신경망 파라미터를 이용하는 방법을 제공할 수 있다.

[0031] 본 개시에 의하면, 압축 파라미터를 기반으로 신경망 파라미터를 학습하는 방법을 제공할 수 있다.

[0032] 본 개시에 의하면, 일반적인 코덱을 이용하여 부호화/복호화할 수 있도록, 특징맵 잠재 표현의 포맷을 변환/역변환하는 방법을 제공할 수 있다.

[0033] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0034] 도 1은 인공신경망의 하나인 Fast R-CNN을 이용하여 물체 검출 및 분류를 수행한 기계 작업 결과의 예시이다.

도 2는 다계층 특징맵을 예시한 도면이다.

도 3은 Mask-RCNN 구조의 인경신경망 모델을 타나낸다.

도 4는 Mask R-CNN의 FPN을 통해 추출된 다계층 특징맵 P_k를 예시한 것이다.

도 5는, YOLO v3를 기반으로 다계층 특징맵이 추출되는 예를 나타낸 것이다.

도 6은 특징맵 추출부와 작업 수행부가 상이한 장치에 존재하는 예를 나타낸 것이다.

도 7은 본 개시에 따른 신경망 기반 영상 부호화/복호화 방법의 개략도이다.

도 8은 신경망 기반의 다계층 특징맵 부호화/복호화 과정을 나타낸 도면이다.

도 9는 이득 유닛 및 역이득 유닛이 포함된 신경망 기반의 다계층 특징맵 부호화/복호화 과정을 나타낸 도면이다.

도 10은 본 개시의 일 실시예에 따른, 영상 부호화기 및 영상 복호화기의 구성도이다.

도 11은 신경망 기반의 다계층 특징맵 부호화 모델을 학습하기 위한 시스템 구성도를 나타낸 것이다.

도 12는 특징맵 잠재 표현 추출부 및 특징맵 복원부의 학습 방법을 설명하기 위한 것이다.

도 13은 미세 조정 가능한 특징맵 복원부를 포함하는 영상 복호화기의 구성도를 나타낸 것이다.

도 14는 본 개시의 일 실시예에 따른 미세 조정 가능한 특징맵 복원부의 구성을 나타낸 것이다.

도 15 및 도 16은 각각 본 개시에 따른 특징맵 부호화 방법 및 특징맵 복호화 방법의 흐름도이다.

도 17은 특징맵 잠재 표현 추출부에서 순차적 융합을 통해 특징맵 잠재 표현을 추출하는 예를 나타낸 것이다.

도 18 내지 도 23은 본 개시에 따른 압축 파라미터 무관 특징맵 복원부 및 압축 파라미터 유관 특징맵 복원부를 포함하는 미세 조정이 가능한 특징맵 복원부의 구성도를 나타낸 것이다.

도 24는 원-핫 인코딩 방식으로 압축 파라미터를 변환한 실시예를 나타낸다.

도 25는 본 개시의 일 실시예에 따른 채널 적응부를 나타낸 것이다.

도 26 및 도 27은 미세 조정 가능한 특징맵 복원부가 영상 복호화부에서 이용 가능한 압축 파라미터들에 대해 미세 조정 학습되는 예를 나타낸 것이다.

도 28은 일반적인 특징맵의 구성을 예시한 것이다.

도 29 내지 도 31은, 각각 공간적 배열 방법, 시간적 배열 방법 및 시공간적 배열 방법을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0035] 본 개시는 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 개시를 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 개시의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다. 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다. 후술하는 예시적 실시예들에 대한 상세한 설명은, 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 실시예를 실시할 수 있기에 충분하도록 상세히 설명된다. 다양한 실시예들은 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 개시의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 실시예의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 예시적 실시예들의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다.

[0036] 본 개시에서 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 개시의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0037] 본 개시의 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있으나, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결 되어"있다거나 "직접 접속되어"있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0038] 본 개시의 실시예에 나타나는 구성부들은 서로 다른 특징적인 기능들을 나타내기 위해 독립적으로 도시되는 것으로, 각 구성부들이 분리된 하드웨어나 하나의 소프트웨어 구성단위로 이루어짐을 의미하지 않는다. 즉, 각 구성부는 설명의 편의상 각각의 구성부로 나열하여 포함한 것으로 각 구성부 중 적어도 두개의 구성부가 합쳐져 하나의 구성부로 이루어지거나, 하나의 구성부가 복수 개의 구성부로 나뉘어져 기능을 수행할 수 있고 이러한 각 구성부의 통합된 실시예 및 분리된 실시예도 본 개시의 본질에서 벗어나지 않는 한 본 개시의 권리범위에 포함된다.

[0039] 본 개시에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 개시를 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 개시에서, "포함하다" 또는 "가지다" 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. 즉, 본 개시에서 특정 구성을 "포함"한다고 기술하는 내용은 해당 구성 이외의 구성을 배제하는 것이 아니며, 추가적인 구성이 본 개시의 실시 또는 본 개시의 기술적 사상의 범위에 포함될 수 있음을 의미한다.

[0040] 본 개시의 일부의 구성 요소는 본 개시에서 본질적인 기능을 수행하는 필수적인 구성 요소는 아니고 단지 성능을 향상시키기 위한 선택적 구성 요소일 수 있다. 본 개시는 단지 성능 향상을 위해 사용되는 구성 요소를 제외한 본 개시의 본질을 구현하는데 필수적인 구성부만을 포함하여 구현될 수 있고, 단지 성능 향상을 위해 사용되는 선택적 구성 요소를 제외한 필수 구성 요소만을 포함한 구조도 본 개시의 권리범위에 포함된다.

[0041] 이하, 도면을 참조하여 본 개시의 실시 형태에 대하여 구체적으로 설명한다. 본 명세서의 실시예를 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 명세서의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략하고, 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

[0042] 인공신경망(ANN, Artificial Neural Network)을 이용한 영상 처리(Image Processing) 기반의 기계 작업(Machine Tack) 활용 사례가 증가 추세에 있다. 일 예로, 입력 영상으로부터, 물체 분류(Object Classification), 물체 인식(Object Recognition), 물체 검출(Object Detection), 물체 영역 분할(Object

Segmentation) 또는 물체 추적(Object Tracking) 등과 같은 기계비전(Machine Vision) 작업 또는, 입력 영상에 대해 해상도 향상(Super-Resolution) 또는 프레임 보간(Frame-Interpolation)등 영상 처리(Image Processing) 작업 등의 활용도가 증가하고 있다.

[0043] 도 1은 인공신경망의 하나인 Fast R-CNN을 이용하여 물체 검출 및 분류를 수행한 기계 작업 결과의 예시이다.

[0044] 위와 같은 기계 작업을 위해, 인간 시각이 아닌 기계 시각을 위한 영상 압축 기술의 필요성이 대두되고 있다.

[0045] 기계 시각을 위한 영상 압축 기술은, 압축 비트량을 최소화하지만, 인간 시각을 위한 영상 압축 기술과 달리, 복원 영상의 화질이 아닌 복원 특징맵을 통한 기계 시각 작업의 성능을 최대화하는 것을 목표로 한다.

[0046] 한편, 기계 작업을 수행하는 인공신경망 모델은, 입력 데이터 또는 입력 영상으로부터, 특징을 추출하는 특징맵 추출부 및 추출된 특징에 기반하여 실제 기계 작업을 수행하는 작업 수행부를 포함할 수 있다.

[0047] 인공신경망 모델에 입력되는 데이터가 영상 형태인 경우, 특징맵 추출부로부터 추출되는 특징을, 특징맵이라 호칭할 수 있다. 이에, 본 개시에서는, 특징맵 추출부로부터 추출되는 특징을 '특징맵'이라 호칭하기로 한다. 다만, 추출되는 특징이 맵(Map) 형태가 아닌 경우에도, 본 개시에서 설명하는 실시예들이 적용될 수 있다 할 것이다.

[0048] 본 개시에서 설명하는 실시예들은, 다계층 특징맵에 적용될 수 있다.

[0049] 도 2는 다계층 특징맵을 예시한 도면이다.

[0050] 다계층 특징맵은, 서로 다른 해상도를 갖는 특징맵들이 복수의 계층을 이루는 구조를 갖는다. 특징맵이 높은 계층에 속할 수록 특징맵의 해상도는 낮아지며, 낮은 계층에 속할 수록 특징맵의 해상도는 높아진다. 이에 따라, 다계층 특징맵은, 피라미드 구조 특징맵이라 호칭할 수도 있다. 한편, 동일 계층 내의 특징맵들의 해상도는, 상호 동일할 수 있다.

[0051] 도 3은 Mask-RCNN 구조의 인경신경망 모델을 타나낸다.

[0052] 도 3에 도시된 Mask-RCNN 구조의 인공 신경망 모델은, 주로 물체 영역 분할 기계 작업을 위해 활용될 수 있다.

[0053] 도 3에 도시된 예에서, 특징 피라미드 네트워크(FPN: Feature Pyramid Network)가 다계층 특징맵 추출부에 해당하고, 영역 제안 네트워크(RPN: Region Proposal Network) 및 관심영역 헤드(ROI Heads: Region Of Interest Heads)가 기계 작업 수행부에 해당할 수 있다.

[0054] 도 3에 도시된 예에서, 특징 파라미드 네트워크(FPN)는 C계층 특징맵 및 P계층 특징맵을 추출하는 것으로 예시하였다. 여기서, C계층 특징맵 및 P계층 특징맵 각각은, 다계층 특징맵일 수 있다.

[0055] 후술되는 실시예들은, P계층 특징맵에 중심으로 설명될 것이나, P계층 특징맵과 다른 형태의 특징맵에 대해서도, 본 개시에서 설명하는 실시예들이 동일하게 적용될 수 있다. 한편, 다계층 특징맵 뿐만 아니라, 단일 계층 특징맵에 대해서도, 본 개시에서 설명하는 실시예들이 동일하게 적용될 수 있다.

[0056] 특징맵은, 2차원 배열로 나타내어질 수 있다. 이에 따라, 특징맵의 크기는, (가로x세로)로 정의될 수 있다.

[0057] 한편, 임의 계층에 속하는 특징맵은, 하나 이상의 채널로 구성될 수 있다. 이에 따라, 각 계층의 특징맵은, (가로x세로x채널 개수)의 크기를 갖는 3차원 배열로 표현될 수 있다.

[0058] 즉, 계층 k에 속하는 특징맵을 F_k라 할 경우, 특징맵 F_k는 추출된 특징값들로 구성된 3차원 배열 $F_k[x][y][c]$로 나타낼 수 있다. 이때, x와 y는 각각 특징값의 수평 방향 위치 및 수직 방향 위치를 나타내고, c는 채널 인덱스를 나타낸다.

[0059] 예를 들어, FPN에서 추출된 다계층 특징맵 C_k 또는 다계층 특징맵 P_k가, 본 발명의 다계층 특징맵 F_k일 수 있다.

[0060] 도 4는 Mask R-CNN의 FPN을 통해 추출된 다계층 특징맵 P_k를 예시한 것이다.

[0061] 도 4에서는, 각 계층에 속하는 특징맵의 첫번째 채널에 속하는 특징맵만이 예시되어 있다.

[0062] FPN에서 추출된 다계층 특징맵 P_k에서, 상이한 계층에 속하는 특징맵들의 해상도는 서로 상이할 수 있다. 일 예로, 도 4에 도시된 예에서와 같이, 계층이 깊어질수록, 특징맵의 너비 및 높이는 더 작아질 수 있다.

[0063] 반면, 상이한 계층에 속하는 특징맵들이라 하더라도, 채널수는 상호 동일할 수 있다. 일 예로, 도 4에 도시된

예에서와 같이, 각 계층의 특징맵은, 256개의 채널로 구성되어 있을 수 있다.

[0064] FPN, 즉, 특징맵 추출부는, 5개 계층의 특징맵, 즉, P_2 내지 P_6를 추출할 수 있다.

[0065] 한편, 추출된 특징맵을 부호화하는 경우, 추출된 5개 계층의 특징맵을 모두 부호화할 수 있다.

[0066] 또는, 해상도가 가장 작은 계층이 부호화/복호화는 생략하고, 인접 계층을 이용하여, 부호화/복호화가 생략된 계층을 복원할 수도 있다.

[0067] 일 예로, 도 4에 도시된 예에서, 4개 계층, 예컨대, P_2 부터 P_5 까지의 특징맵들만을 부호화하고, 부호화가 생략된 계층은, 명시적으로 부호화/복호화되는 계층의 특징맵으로부터 유도할 수 있다. 일 예로, P_6 특징맵의 부호화/복호화가 생략된 경우, 복호화된 P_5 특징맵을 기반으로, P_6 특쟁맵을 복원할 수 있다.

[0068] YOLO v3와 같은 기계 작업을 위한 모델에서도 FPN의 P 계층 특징맵과 유사한 방법으로 다계층 특징맵을 추출하고, 이를, 기계 작업 수행에 이용할 수 있다.

[0069] 도 5는, YOLO v3를 기반으로 다계층 특징맵이 추출되는 예를 나타낸 것이다.

[0070] 도 5에서는, 세 개의 계층들(즉, (Output1, Output2, Output3)로 이루어진 다계층 특징맵이 추출되고, 이를 기반으로 기계 작업이 수행되는 것으로 예시되었다.

[0071] 구체적으로, 도 5에서는, YOLO v3 내에서, FPN과 유사한 피라미드 구조를 사용하는, darknet 53이 다계층 특징맵 추출부로서 기능하는 것으로 예시되었다.

[0072] 도 4를 통해 설명한 바와 마찬가지로, 추출된 특징맵을 부호화함에 있어서, 3개의 계층들을 모두 부호화/복호화할 수 있다.

[0073] 또는, 추출된 특징맵들 중 해상도가 가장 작은 계층의 특징맵에 대해서는, 부호화/복호화를 생략할 수 있다. 이 경우, 명시적으로 부호화/복호화되는 다른 계층의 특징맵을 기반으로, 부호화/복호화가 생략된 계층의 특징맵을 복원할 수 있다.

[0074] 한편, 기계 작업이 대형 서버 뿐 아니라, 모바일 장치와 같이 컴퓨팅 능력이 상대적으로 낮은 장치에까지 사용됨에 따라, 특징맵 추출부와, 작업 수행부가 같은 장치 내에 존재하지 않는 경우도 점차 많아지고 있다.

[0075] 도 6은 특징맵 추출부와 작업 수행부가 상이한 장치에 존재하는 예를 나타낸 것이다.

[0076] 구체적으로, 도 6에서는, 특징맵 추출부는, 모바일 장치에 존재하는 반면, 물체 분할, 디스패러티맵 추정, 또는 영상복원 등과 같은 기계 작업을 수행하기 위한 작업 수행부는, 클라우드 서버에 존재하는 것으로 예시되었다.

[0077] 이 경우, 모바일 장치에서 추출된 특징맵이 서버로 전달되고, 서버에서의 작업 수행 결과가 모바일 장치로 피드백될 수 있다.

[0078] 도 6에 도시된 예에서와 같이, 특징맵 추출부 및 작업 수행부가 분리되어 있는 경우, 추출된 특징맵을, 작업 수행부에 전달해야 한다. 이에 따라, 전달 및 저장될 특징맵의 데이터량을 최소화하는 동시에, 작업 수행 성능의 저하를 최소화하기 위한 특징맵 부호화/복호화 방법이 요구될 수 있다.

[0079] 특징맵 추출부와 작업 수행부가 하나의 장치에 존재하는 경우라 하더라도, 추출된 특징맵을 저장해 두었다가 추후 작업 수행부에서 활용하는 경우가 있을 수 있다. 이 경우에도, 저장될 특징맵의 데이터량을 최소화하기 위한 특징맵 부호화/복호화 방법이 요구될 수 있다.

[0080] 이에, 본 개시에서는, 특징맵 추출부에서 추출된 특징맵, 구체적으로는, 특징맵을 부호화/복호화하는 방법을 제안한다.

[0081] 본 개시에서, '영상'은, 카메라를 통해 획득된 자연 영상 뿐만 아니라, 컴퓨터 그래픽스, 홀로그램 영상, 신경망을 통해 추출된 특징맵 영상 또는 초음파 영상 등 다양한 종류의 영상을 나타낼 수 있다.

[0082] 도 7은 본 개시에 따른 신경망 기반 영상 부호화/복호화 방법의 개략도이다.

[0083] 본 개시에 따른, 신경망 기반 영상 부호화/복호화 방법은, 부호화 신경망, 복호화 신경망, 양자화, 잠재표현 확률 모형, 엔트로피 부호화, 엔트로피 복호화 등을 통해 구현될 수 있다.

[0084] 먼저, 부호화기에서는, 부호화 신경망을 통해, 부호화하려는 영상(즉, 입력 영상) x를 잠재 표현(Latent

representation) y로 변환할 수 있다.

[0085] 여기서, 잠재 표현은, 은닉(잠재) 벡터(Latent vector), 잠재 표현(표상)(Latent representation) 또는 은닉(잠재) 특징맵 (Latent feature map) 중 적어도 하나를 의미할 수 있다.

[0086] 부호화기에서는, 잠재 표현 y의 각 성분 y_i를 양자화(Quantization)할 수 있다.

[0087] 양자화된 잠재 표현 $\hat{y}$는, 학습 가능한 잠재 표현 확률 모형 $p_{\hat{y}}(\hat{y})$ 에 기초한 엔트로피 부호화를 통해 비트스트림(Bitstream) 형태로 변환되어, 복호화기로 전송될 수 있다.

[0088] 복호화기에서는, 비트스트림을 수신하고, 부호화기에서 동일한 잠재 표현 확률 모형에 기초한 엔트로피 복호화를 통해, 양자화된 잠재 표현 $\hat{y}$를 복원할 수 있다.

[0089] 또한, 복호화 신경망은, 양자화된 잠재 표현 $\hat{y}$의 입력에 대한 응답으로, 복원 영상 $\hat{x}$를 출력할 수 있다.

[0090] 한편, 도 7에 도시된 예에서, 점선으로 표현된 블록은, 학습 가능한 파라미터 또는 학습 가능한 신경망임을 나타낸다.

[0091] 도 7에 도시된 학습 가능한 블록들의 신경망 파라미터는, 손실함수 및 역전파 알고리즘(Backpropagation algorithm)을 통해 학습될 수 있다.

[0092] 한편, 다계층 특징맵은, 공간적 해상도가 상이한 여러 개의 특징맵들로 구성되어 있다. 이에 따라, 다계층 특징맵을 부호화/복호화하는 데에는, 자연 영상을 부호화/복호화할 때와 다른 방법이 요구될 수 있다.

[0093] 이에, 본 개시에서는, 입력 영상이 다계층 특징맵 $\{F_k\}$ (k는, 1 내지 L, L은 계층들의 개수)인 경우, 신경망 기반의 영상 부호화/복호화 방법을 제시한다. 한편, 변수 k의 값이 작을수록 공간적 해상도가 큰 계층의 특징맵을 의미할 수 있다.

[0094] 도 8은 신경망 기반의 다계층 특징맵 부호화/복호화 과정을 나타낸 도면이다.

[0095] 부호화기에서는, 부호화하려는 다계층 특징맵 $\{F_k\}$을 부호화 신경망을 통해 잠재 표현 y로 변환할 수 있다.

[0096] 또한, 부호화기에서는, 잠재 표현 y의 각 성분 y_i를 양자화(Quantization)할 수 있다.

[0097] 양자화된 잠재 표현 $\hat{y}$는, 학습 가능한 잠재 표현 확률 모형 $p_{\hat{y}}(\hat{y})$에 기초한 엔트로피 부호화를 통해 비트스트림(Bitstream) 형태로 변환되어, 복호화기로 전송될 수 있다.

[0098] 복호화기에서는, 비트스트림을 수신하고, 부호화기에서 동일한 잠재 표현 확률 모형에 기초한 엔트로피 복호화를 통해, 양자화된 잠재 표현 $\hat{y}$를 복원할 수 있다.

[0099] 또한, 복호화 신경망은, 양자화된 잠재 표현 $\hat{y}$의 입력에 대한 응답으로, 복원된 다계층 특징맵 $\{\widehat{F_k}\}$를 생성한다.

[0100] 한편, 도 8에 도시된 예에서, 점선으로 표현된 블록은, 학습 가능한 파라미터 또는 학습 가능한 신경망임을 나타낸다.

[0101] 도 8에 도시된 학습 가능한 블록들의 신경망 파라미터는, 손실함수 및 역전파 알고리즘(Backpropagation algorithm)을 통해 학습될 수 있다.

[0102] 한편, 신경망 기반 영상 부호화/복호화시, 가변 비트율 부호화/복호화를 지원하기 위해, 이득 유닛(GU, Gain Unit) 및 역이득 유닛(IGU, Inverse Gain Unit)을 이용할 수 있다.

[0103] 도 9는 이득 유닛 및 역이득 유닛이 포함된 신경망 기반의 다계층 특징맵 부호화/복호화 과정을 나타낸 도면이다.

[0104] 도 9에 도시된 예에서와 같이, 이득 유닛은, 부호화 신경망의 끝에 위치할 수 있다. 반면, 역이득 유닛은, 복호화 신경망의 시작부에 위치할 수 있다.

[0105] 이득 유닛은, 가변 비트율 부호화를 위한 Q개의 이득 벡터 $\{v_q\}$ (GV: Gain Vector, q는, 1 내지 Q)들 중 하나를

이용하여 입력 특징맵의 각 채널을 스케일링 할 수 있다. 여기서, 정수 q는, 비트율 수준을 나타내고, Q는, 가변 비트율 부호화를 위한 비트율 수준의 개수를 나타낸다. q는, 0보다 큰 값을 가질 수 있다.

[0106] 이득 유닛은, 이득 유닛을 거치기 직전의 잠재 표현 (이하, 중간 잠재 표현 τ라 함)을 다음의 수학식 1에 따라 스케일링할 수 있다.

수학식 1

[0107] $$y[c][h][w]=\tau[c][h][w]\times v_q[c]$$

[0108] 상기 수학식 1에서, c (1부터 C 까지의 값을 가짐)는, 중간 잠재 표현 τ의 채널 인덱스를 나타낸다. 이득 벡터의 길이는 C와 같을 수 있다. h (1 부터 H까지의 값을 가짐) 및 w (1부터 W까지의 가짐)는 각각, 중간 잠재 표현 τ의 수직 방향 좌표 및 수평 방향 좌표를 나타낸다. H 및 W는, 특징 맵의 높이 및 너비를 나타낼 수 있다.

[0109] 역이득 유닛은, 이득 벡터에 대응하는 역이득 벡터들 $\{u_q\}$ (q는 1 내지 Q) 중 하나를 이용하여, 입력 특징맵의 각 채널들을 스케일링할 수 있다.

[0110] 역이득 유닛에서 출력되는 중간 잠재 표현 η는 다음 수학식 2와 같은 스케일링 과정을 통해 획득될 수 있다.

수학식 2

[0111] $$\eta[c][h][w]=\hat{y}[c][h][w]\times u_q[c]$$

[0112] 상기 수학식 2에서, c (1부터 C 까지의 값을 가짐)는, 양자화된 잠재 표현 $\hat{y}$의 채널 인덱스를 나타낸다. 역이득 벡터의 길이는 C와 같을 수 있다. h (1 부터 H까지의 값을 가짐) 및 w (1부터 W까지의 가짐)는 각각, 중간 잠재 표현 $\hat{y}$의 수직 방향 좌표 및 수평 방향 좌표를 나타낸다. H 및 W는, 특징 맵의 높이 및 너비를 나타낼 수 있다.

[0113] 이득 벡터 및 역이득 벡터 쌍(v_q, u_q)의 각 성분값은, 주어진 비트율 수준 q에 대응하는 비트율 제약 조건을 만족하도록, 학습을 통해 최적화될 수 있다.

[0114] 결과적으로, 현재 부호화할 영상의 비트율 및 복원 화질은, 이득 유닛 및 역이득 유닛에게 주어지는 비트율 수준 q에 따라 결정될 수 있다.

[0115] 상술한 바와 같이, 신경망 기반의 다계층 특징맵의 부호화는, 학습 가능한 파라미터로 구성된 신경망들을 통해 구성될 수 있다. 위와 같은 신경망들을 포함하는 부호화 모델을, 신경망 기반의 다계층 특징맵 부호화 모델이라 호칭할 수 있다.

[0116] 한편, 신경망의 학습은, 역전파 알고리즘(Backpropagation Algorithm)에 기초할 수 있다. 역전파 알고리즘을 통해, 학습 데이터로부터 계산된 특정 손실함수를 최소화하는 방향으로 파라미터와 신경망의 가중치들을 갱신할 수 있다.

[0117] 신경망 기반의 다계층 특징맵 부호화 모델을 학습할 때, 율-왜곡(Rate-distortion) 최적화 방법 또는 율-성능(Rate-performance) 최적화 방법을 사용할 수 있다.

[0118] 율-왜곡 최적화 방법은, 입력되는 다계층 특징맵과 복원된 다계층 특징맵 간의 왜곡(Distortion)과 부호화기에서 복호화기로 전송하는 비트스트림의 비트율(Bit-rate)을 동시에 최소화하도록 학습을 진행하는 것이다.

[0119] 율-성능 최적화 방법은, 복원된 다계층 특징맵을 통해 수행된 기계 작업 수행 성능(Performance)과 부호화기에서 복호화기로 전송하는 비트스트림의 비트율을 동시에 최소화하도록 학습을 진행하는 것이다.

[0120] 신경망 기반의 다계층 특징맵 부호화 모델에서 사용하는 왜곡 손실함수는, 수학식 3에서와 같이, 입력 다계층 특징맵 $\{F_k\}$과 복원된 다계층 특징맵 $\{\widehat{F_k}\}$간의 평균 제곱 오차 (MSE: Mean Square Error) 또는 MS-SSIM (Multi-Scale Structural Similarity Index Measure)를 가중합하여 획득될 수 있다.

[0121] 한편, 가중합 연산시, 계층 별로 가중치 w_k가 설정될 수 있다.

수학식 3

[0122] $$L_D = E_{x \sim p_x}\left(\sum_{k=1,\ldots,L} w_k \times D(F_k, \widehat{F_k}) \right)$$

[0123] 상기 수학식 3에서, D는, MSE 또는 MS-SSIM 과 같은 왜곡 함수를 나타낸다.

[0124] 비트율은, 잠재표현 확률 모형에 의해 추정된 잠재표현의 확률 분포와 실제 발생한 잠재표현 간의 교차 엔트로피 (Cross-entropy)로 근사할 수 있다. 수학식 4는, 비트율을 산출하는 예를 나타낸 것이다.

수학식 4

[0125] $$L_R = E_{x \sim p_x}\left(-\log p_{\tilde{y}}(\tilde{y})\right)$$

[0126] 기계 작업 손실함수 L_p는 압축 및 복원된 다계층 특징맵으로부터 수행된 기계 작업의 성능을 의미한다. 기계 작업 성능은, 정답 레이블(Label)과 기계 작업의 추론 결과 사이의 비교를 통해 계산될 수 있다.

[0127] 이때, 기계 작업의 종류에 따라 분류 손실 함수(Classification loss), 바운딩 박스 손실함수 (bounding box loss) 또는 마스크 손실 함수 (mask loss) 중 적어도 하나가 사용될 수 있다. 다만, 열거된 것과 다른 종류의 손실 함수를 기반으로, 본 개시에 따른 실시예들이 수행될 수도 있다.

[0128] 율-왜곡 최적화를 위한 손실함수 L_{RD}은 왜곡 손실함수 L_D와 교차 엔트로피 기반 손실함수 L_R 을 기반으로 획득될 수 있다. 일 예로, 수학식 5는, 손실함수 L_{RD}를 유도하는 예를 나타낸다.

수학식 5

[0129] $$L_{RD} = L_R + \lambda \times L_D$$

[0130] 한편, 손실함수 L_{RD}를 유도함에 있어서, 왜곡 손실함수 L_D와 교차 엔트로피 기반 손실함수 L_R 사이의 비율을 조절하기 위한 변수 λ가 이용될 수 있다. 변수 λ에 기초하여, 원하는 복원 품질 수준 및 비트율을 결정할 수 있다. 일 예로, 변수 λ의 값이 클수록, 복원 품질 수준이 높다.

[0131] 율-성능 최적화를 위한 손실함수 L_{RP}는, 수학식 6에 따라 획득될 수 있다.

수학식 6

[0132] $$L_{RP} = L_R + \lambda \times L_P$$

[0133] 수학식 6의 예에서와 같이, 손실함수 L_{RP}는, 성능 손실함수 L_P 및 교차 엔트로피 기반 손실함수 L_R을 기반으로 획득될 수 있다.

[0134] 한편, 손실함수 L_{RP}를 유도함에 있어서, 성능 손실함수 L_P와 교차 엔트로피 기반 손실함수 L_R 사이의 비율을 조절하기 위한 변수 λ가 이용될 수 있다. 변수 λ에 기초하여, 원하는 복원 성능 수준 및 비트율을 결정할 수 있다. 일 예로, 변수 λ의 값이 클수록, 기계 작업 성능이 높다.

[0135] 도 10은 본 개시의 일 실시예에 따른, 영상 부호화기 및 영상 복호화기의 구성도이다.

[0136] 특징맵 추출부 및 작업 수행부가 별도의 장치로서 구현되는 경우, 영상 부호화기는, 특징맵 추출부를 포함하는

장치를 나타내고, 영상 복호화기는, 작업 수행부를 포함하는 장치를 나타낼 수 있다.

[0137] 또는, 특징맵 추출부 및 작업 수행부가 하나의 장치로서 구현되는 경우, 영상 부호화기 및 영상 복호화기는 동일한 장치일 수 있다.

[0138] 도 10을 참조하면, 영상 부호화기(10)는, 특징맵 부호화부(110) 및 내부 부호화부(120)를 포함하고, 영상 복호화기(20)는, 내부 복호화부(210) 및 특징맵 복호화부(220)를 포함할 수 있다.

[0139] 한편, 내부 부호화부(120)는, 포맷 변환부(130) 및 영상 부호화부(140)를 포함하고, 내부 복호화부(210)는, 영상 복호화부(230) 및 포맷 역변환부(240)를 포함할 수 있다.

[0140] 또한, 추출된 특징맵으로부터 도출된 영상을 실제로 부호화/복호화하는 내부 부호화부(120) 및 내부 복호화부(210)를, 내부 코덱(Inner Codec)이라 호칭할 수도 있다.

[0141] 영상 부호화기의 특징맵 부호화부(110)는, 입력되는 특징맵으로부터, 특징맵 잠재 표현 y를 추출할 수 있다. 구체적으로, 특징맵 부호화부(110) 내 특징맵 잠재표현 추출부(112)를 통해, 상기 동작이 수행될 수 있다.

[0142] 특징맵 부호화부(110)로 입력되는 특징맵은, 단일 계층 특징맵 또는 다계층 특징맵일 수 있다.

[0143] 상술한 바와 같이, 특징맵 잠재 표현은, 은닉(잠재) 벡터(Latent vector), 잠재 표현(표상)(Latent representation) 또는 은닉(잠재) 특징맵 (Latent feature map) 중 하나를 의미할 수 있다.

[0144] 내부 부호화부(120) 내 포맷 변환부(130)에서는, 잠재 표현 y를 포맷 변환된 잠재 표현 t로 변환한다. 포맷 변환부(130)는, 특징맵들을 배열하기 위한 특징맵 배열부 및 특징값들의 양자화를 위한 양자화부를 포함할 수 있다.

[0145] 내부 부호화부(120) 내 영상 부호화부(140)에서는, 포맷 변환된 잠재 표현 t를 부호화하여, 비트스트림을 생성한다.

[0146] 내부 복호화부(210) 내 영상 복호화부(230)에서는, 비트스트림을 복호화하여, 포맷 변환된 잠재 표현 $\hat{t}$를 생성한다.

[0147] 내부 복호화부(210) 내 포맷 역변환부(240)는, 포맷 변환된 잠재 표현 $\hat{t}$에 대한 포맷 역변환을 수행하여, 복원된 잠재 표현 $\hat{y}$를 생성한다. 포맷 역변환을 수행함으로써, 복원된 잠재 표현 $\hat{y}$가 원래 잠재 표현 t와 동일한 포맷을 갖게 될 수 있다. 포맷 역변환부(240)는, 배열 변환된 특징맵들을 원래 배열로 복원하기 위한 특징맵 재배열부 및 양자화된 특징값들의 역양자화를 위한 역양자화부를 포함할 수 있다.

[0148] 특징맵 복호화부(220)는, 복원된 잠재 표현 $\hat{y}$를 특징맵으로 복원할 수 있다. 구체적으로, 특징맵 복호화부(220) 내 특징맵 복원부(222)를 통해, 상기 동작이 수행될 수 있다.

[0149] 한편, 내부 부호화부(120) 및 내부 복호화부(210)에서 주로 이용되는 표준 압축 코덱(예컨대, HEVC 또는 VVC 등)은, 미분이 불가능한 구조이다. 이에 따라, 표준 압축 코덱이 적용된 경우, 종단간 학습은 불가능하다. 이를 고려하여, 내부 코덱(즉, 내부 부호화부(120) 및 내부 복호화부(210))를 제거한 뒤, 특징맵 잠재표현 추출부(112)에서 출력되는 특징맵 잠재 표현을 그대로 특징맵 복원부(222)의 입력으로 사용하여, 신경망 기반의 다계층 특징맵 부호화 모델을 학습할 수 있다.

[0150] 도 11은 신경망 기반의 다계층 특징맵 부호화 모델을 학습하기 위한 시스템 구성도를 나타낸 것이다.

[0151] 도 11에서는, 도 10에 도시된 영상 부호화기(10) 및 영상 복호화기(20)에서, 내부 코덱이 제거된 구성을 기반으로, 학습이 수행되는 것으로 도시되었다.

[0152] 한편, 도 11에 도시된 예에서, 내부 코덱이 제거된 경우, 영상 압축시 발생하는 압축 잡음을 학습할 수 없게 된다. 이에 따라, 왜곡 손실함수 L_D 또는 기계 작업 손실 함수 L_P를 사용하여, 신경망 기반의 다계층 특징맵 부호화 모델을 학습할 수 있다.

[0153] 다만, 왜곡 손실함수 또는 기계 작업 손실함수 만을 사용하여 다계층 특징맵 부호화 모델을 학습하는 경우, 압축율을 고려할 수 없게 된다. 특징맵 잠재표현 추출부(112) 및 특징맵 복원부(222)가 내부 코덱에서의 압축 잡음을 고려하여 학습되도록, 수학식 5의 율-왜곡 손실 함수(rate-distortion loss) 또는 수학식 6의 율-기계 작업 손실함수 중 적어도 하나를 이용하여, 학습 방법을 최적화할 수 있다.

[0154] 이하, 율-왜곡 손실 함수 또는 율-기계 작업 손실 함수 중 적어도 하나 이상을 사용하는 학습 방법에 대해 상세히 설명하기로 한다.

[0155] 도 12는 신경망 기반의 내부 코덱을 사용한 특징맵 잠재 표현 추출부 및 특징맵 복원부의 학습 방법을을 설명하기 위한 것이다.

[0156] 도 12에 도시된 예에 따라, 내부 부호화부(120)는, 신경망 파라미터 학습을 위한 양자화부(160) 및 학습 가능한 엔트로피 부호화부(150)를 더 포함할 수도 있고, 내부 복호화부(130)는, 신경망 파리터 학습을 위한 학습 가능한 엔트로피 복호화부(250) 및 역양자화부(260)를 더 포함할 수도 있다.

[0157] 설명의 편의를 위해, 본 개시의 일부 실시예들에서는, 내부 코덱의 구성에 따라, 내부 코덱을 제1 내부 코덱 또는 제2 내부 코덱으로 구분하여 명명하였다. 일 예로, 제1 내부 코덱은, 학습 가능한 엔트로피 부호화부(150) 및 엔트로피 복호화부(250)를 포함하는 것이고, 제2 내부 코덱은 기계 임무를 수행하기 위한 영상 부호화부 및 영상 복호화부를 포함하는 것일 수 있다.

[0158] 본 개시에서, '제1' 또는 '제2' 와 같은 첨두어가 부가되지 않더라도, 해당 내부 코덱의 이용 목적/용도에 따라, 내부 코덱이 제1 내부 코덱인지 또는 제2 내부 코덱인지 명확히 이해될 수 있을 것이다.

[0159] 도 12에서, 양자화부(160)는, 융합 잠재표현 y의 모든 성분 y_i에 대해 균등 양자화를 수행할 수 있다. 양자화부(160)에서 균등 양자화를 수행하는 경우, 통상의 반올림 연산이 이용될 수 있으며, 역전파 알고리즘을 사용할 수 있도록, 양자화를 균등 노이즈(Uniform Noise)를 더하는 과정으로 대체할 수도 있다. 균등 노이즈는 -0.5와 0.5 사이의 값일 수 있다. 양자화부(160)에서 균등 양자화를 수행하는 경우, 역양자화부(240)는 생략될 수 있다. 양자화의 수행 결과, 양자화된 특징맵 잠재표현이 출력될 수 있다.

[0160] 특징맵 잠재표현 추출부(112) 및 특징맵 복원부(222)를 최적화하기 위해, 학습 가능한 특징맵 잠재 표현의 엔트로피 부호화부(150) 및 학습 가능한 특징맵 잠재 표현의 엔트로피 복호화부(250)이 포함된 신경망 기반의 제 1 내부 코덱을 이용할 수 있다. 즉, 실제 부호화/복호화를 위한 비신경망 기반의 제 2 내부 코덱(즉, 영상 부호화부(140) 및 영상 복호화부(230))과, 학습을 위한 제 1 내부 코덱의 구성(즉, 학습 가능한 엔트로피 부호화부(250) 및 학습 가능한 엔트로피 복호화부(150))이 상이할 수 있다.

[0161] 신경망 기반의 내부 코덱을 사용함으로써, 내부 코덱과 특징맵 부호화부(110) 및 특징맵 복호화부(220) 전체가 종단간 학습이 가능해지고, 이에 따라, 특징맵 부호화 모델의 학습 시 율-왜곡 손실 함수를 통해 압축률 및 왜곡을 동시에 고려할 수 있게 된다.

[0162] 한편, 학습을 위한 내부 코덱(즉, 제 1 내부 코덱)의 구성과, 실제 영상 부/복호화를 위한 내부 코덱(즉, 제 2 내부 코덱)의 구성이 상이하므로, 위 두 경우 간에 있어서 압축 잡음이 상이할 수 있다. 이에 따라, 특징맵 복원부(222)가, 제 1 내부 코덱의 압축 잡음이 아닌, 제 2 내부 코덱의 압축 잡음을 학습할 수 있도록, 특징맵 복원부(222)를 미세 조정 가능한 특징맵 복원부(224)로 대체할 수 있다.

[0163] 구체적으로, 제 1 내부 코덱의 출력인 $\tilde{y}$와 제 2 내부 코덱의 출력인 $\hat{y}$ 은 서로 상이할 수 있다. 따라서 $\tilde{y}$를 입력 받아 특징맵을 복원하도록 최적화된 특징맵 복원부(222)가 $\hat{y}$을 입력 받아 특징맵을 복원하도록 최적화된 미세 조정 가능한 특징맵 복원부(224)로 변환하는 과정이 필요하다.

[0164] 도 13은 미세 조정 가능한 특징맵 복원부를 포함하는 영상 복호화기의 구성도를 나타낸 것이다.

[0165] 도 13에 도시된 예에서는, 특징맵 복호화부(220)는, 일반적인 특징맵 복원부(222)가 아닌, 미세 조정 가능한 특징맵 복원부(224)를 포함하는 것으로 예시되었다.

[0166] 구체적으로, 본 개시에서는, 특징맵 복원부(222)가 학습한 제 1 내부 코덱으로부터 발생되는 압축 잡음과, 실제 사용되는 제 2 내부 코덱의 압축 잡음의 불일치를 최소화할 수 있도록, 일반적인 특징맵 복원부(222) 대신 미세 조정 가능한 특징맵 복원부(224)를 사용할 수 있다.

[0167] 한편, 본 개시에 따른 미세 조정 가능한 특징맵 복원부(224)는, 일반적인 특징맵 복원부(222)를 미세조정 학습이 가능하도록 개량한 것일 수 있다.

[0168] 이하, 제 2 내부 코덱을 사용한 2단계 학습 방법을 통해, 미세 조정 가능한 특징맵 복원부(224)를 취득하는 방법을 설명하기로 한다.

[0169] 도 12에 도시된 예에서와 같이, 1단계 학습을 통해, 오류 역전파로 학습 가능한 신경망 기반의 제 1 내부 코덱을 사용하여, 특징맵 잠재표현 추출부(112) 및 특징맵 복원부(222)를 최적화할 수 있다. 제1 내부 코덱은 통상의 신경망 기반의 영상 압축 모델에서 사용하는 엔트로피 부호화 및 복호화 방법을 사용할 수 있다. 도 8과 도 9에 도시된 예가 제1 내부 코덱의 예시이며, 구체적으로 다음 참조 문헌을 참조하여 이해할 수 있다.

[0170] He, Dailan, et al. "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[0171] 도 13에 도시된 예는 2단계 학습을 통해, 일반적인 특징맵 복원부(222) 대신 미세 조정 가능한 특징맵 복원부(224)로 변환하는 방법을 나타낸다.

[0172] 구체적으로, 제 1 내부 코덱으로 인한 압축 잡음(즉, $\tilde{y}-y$)은 실제 사용할 제 2 내부 코덱으로 인한 압축 잡음(즉, $\hat{y}-y$)은 서로 상이할 수 있으므로, 1단계 학습을 통해 취득한 특징맵 잠재표현 추출부(112)와 특징맵 복원부(222) 및 제 2 내부 코덱을 사용하여 특징맵 복원부를 미세 조정한다. 즉, 실제 사용되는 영상 부호화부(140) 및 영상 복호화부(230)와, 포맷 역변환부(240)을 거친 특징맵 잠재표현 $\hat{y}$을 입력받고, 입력된 특징맵 잠재 표현과 원본 입력 특징맵 간의 오차를 최소화하거나 또는 최종 기계작업 성능을 최대화하도록 학습될 수 있다.

[0173] 한편, 학습에 사용되는 손실 함수는, MSE 또는 MAE 등과 같이 입력 특징맵의 왜곡을 줄이기 위한 손실, 또는, mAP 또는 MOTA 등과 같이 기계작업 성능을 최대화하는 손실 중 적어도 하나를 고려한 것일 수 있다.

[0174] 도 14는 본 개시의 일 실시예에 따른 미세 조정 가능한 특징맵 복원부의 구성을 나타낸 것이다.

[0175] 도 14를 참조하면, 미세 조정 가능한 특징맵 복원부(224)는, 압축 파라미터 무관 특징맵 복원부(228) 및 압축 파라미터 유관 특징맵 복원부(226)를 포함할 수 있다. 한편, 압축 파라미터 유관 특징맵 복원부(226)를 생략한 채, 압축 파라미터 무관 특징맵 복원부(228)만으로, 미세 조정 가능한 특징맵 복원부(224)를 구현할 수도 있다.

[0176] 압축 파라미터 무관 특징맵 복원부(228)는, 영상 부호화기의 압축 파라미터와 무관하게, 단일한 신경망 가중치 파라미터 세트를 갖도록 학습될 수 있다. 압축 파라미터 무관 특징맵 복원부(228)는 특징맵 복원부(222)와 동등한 구성을 가질 수도 있고, 또는, 특징맵 복원부(222)의 일부 구성요소를 제외하여, 압축 파라미터 무관 특징맵 복원부(228)를 구성할 수도 있다.

[0177] 압축 파라미터 유관 특징맵 복원부(226)는, 영상 부호화기의 압축 파라미터에 적응적인 신경망 가중치 파라메터 세트를 갖도록 학습될 수 있다. 한편, 미세 조정 가능한 특징맵 복원부(224) 내에서, 압축 파라미터 유관 특징맵 복원부(226)는 압축 파라미터 무관 특징맵 복원부(228)보다 앞단에 위치할 수 있다. 즉, 미세 조정 가능한 특징맵 복원부(224)의 입력은, 압축 파라미터 유관 특징맵 복원부(226)로 입력되고, 압축 파라미터 유관 특징맵 복원부(226)의 출력은, 압축 파라미터 무관 특징맵 복원부(228)로 입력될 수 있다.

[0178] 한편, 본 개시에서, 영상 부호화기의 압축 파라미터는, 영상 부호화기의 압축 수준을 결정하는 것일 수 있다. 일 예로, 압축 파라미터는, AVC, HEVC 또는, VVC 등의 코덱에서 사용되는 양자화 매개변수(QP: Quantization Parameter)를 포함할 수 있다.

[0179] 압축 파라미터 유관 특징맵 복원부(226)는 압축 파라미터 적응부(226-2) 또는 채널 적응부(226-1) 중 적어도 하나를 포함할 수 있다.

[0180] 일 예로, 압축 파라미터 유관 특징맵 복원부(226)는, 압축 파라미터 적응부(226-2) 만으로 구성되거나, 압축 파라미터 적응부(226-2)와 채널 적응부(226-1)를 모두 포함하도록 구성될 수도 있다. 또는, 압축 파라미터 유관 특징맵 복원부(226)는, 채널 적응부(226-1)만으로 구성될 수 있다.

[0181] 압축 파라미터 적응부(226-2)는, 특징맵 복원부(222)의 일부 구성 요소를 사용하여 구현될 수도 있고, 또는 특징맵 복원부(222)와는 별개인 구성요소를 사용하여 구현될 수도 있다.

[0182] 상술한 개시를 바탕으로, 본 개시에 따른 특징맵 부호화/복호화 및 학습 방법에 대해 상세히 살펴보기로 한다.

[0183] 도 15 및 도 16은 각각 본 개시에 따른 특징맵 부호화 방법 및 특징맵 복호화 방법의 흐름도이다.

[0184] 도 15를 참조하면, 본 개시에 따른, 영상 부호화 방법은, [E1] 특징맵 잠재 표현을 추출하는 단계, [E2] 특징맵

잠재 표현의 포맷을 변환하는 단계 및 [E3] 포맷 변환된 특징맵 잠재 표현을 부호화하는 단계를 포함할 수 있다.

[0185] 도 16을 참조하면, 본 개시에 따른 영상 복호화 방법은, [D3] 비트스트림으로부터 포맷 변환된 특징맵 잠재 표현을 복호화하는 단계, [D2] 복호화된 포맷 변환된 특징맵 잠재 표현에 대해 포맷 역변환을 수행하는 단계 및 [D1] 역번환된 특징맵 잠재 표현을 특징맵으로 복원하는 단계를 포함할 수 있다.

[0186] 이하, 도 15 및 도 16의 각 단계에 대해 상세히 설명하기로 한다.

[0187] **[E1] 특징맵 잠재 표현을 추출하는 단계**

[0188] 특징맵 부호화부(110), 구체적으로는, 특징맵 잠재 표현 추출부(112)는, 입력되는 특징맵으로부터, 특징맵 잠재 표현을 추출한다.

[0189] 이때, 특징맵 잠재 표현 추출부(112)는, 신경망 기반의 특징맵 잠재표현 엔트로피 부호화부(150) 및 신경망 기반의 특징맵 잠재표현 엔트로피 복호화부(250)를 기반으로 학습된 파라미터들을 사용할 수 있다. 즉, 특징맵 잠재 표현 추출부(112)는, 종단 간 학습을 통해 최적화된 파라미터를 이용하여, 특징맵 잠재 표현을 추출할 수 있다.

[0190] 특징맵 잠재표현 추출부(112)는 입력 특징맵에 패딩(padding)을 수행하는 과정을 포함할 수 있다. 신경망 기반의 특징맵 잠재표현 추출부가 사용되는 경우, 입력 특징맵의 해상도가 신경망 기반의 특징맵 잠재표현 추출부가 수용 가능한 입력과 맞지 않는 경우가 발생할 수 있다. 이 경우, 입력 특징맵에 패딩을 적용할 수 있다.

[0191] 특징맵 잠재표현 추출부(112)에서 패딩이 수행된 경우, 특징맵 복원부(222) 또는 미세 조정 가능한 특징맵 복원부(224)에서 언패딩(unpadding) 과정을 수행할 수 있다. 언패딩 과정을 통해, 복원된 특징맵을, 패딩 전의 특징맵과 동일한 해상도로 변환할 수 있다.

[0192] 특징맵 복원부에서, 언패딩을 수행할 수 있도록, 패딩에 관한 정보가 부호화되어 시그날링될 수 있다.

[0193] 특히, 다계층 특징맵의 특징맵 잠재 표현 추출을 위해서는 연산량 또는 모델 크기를 줄일 필요가 있다. 이에 따라, 도 17에 도시된 예에서와 같은, 순차적 융합 부호화 방법이 이용될 수 있다.

[0194] 도 17은 특징맵 잠재 표현 추출부에서 순차적 융합을 통해 특징맵 잠재 표현을 추출하는 예를 나타낸 것이다.

[0195] 다계층 특징맵인 P 계층 특징맵 P_k (k는 2 부터 5)가 입력되면, 다계층 특징맵 내 각 계층을 순차적으로 특징맵 잠재 표현으로 변환할 수 있다.

[0196] 일 예로, p_2를 입력으로 하는 제1 부호화 블록은, Resblock 및 Resblock 2↓로 구성되고, 중간 잠재표현인 l_1을 출력한다. 여기서, ↓은, 1/2 다운 스케일된 것을 의미한다.

[0197] 제2 부호화 블록은 Resblock 및 Resblock 2↓로 구성되고, l_1 및 p_3의 입력에 대한 응답으로, 융합 잠재표현인 f_2를 출력한다.

[0198] 제3 부호화 블록은 Resblock 및 Resblock 2↓로 구성되고, f_2 및 p_4의 입력에 대한 응답으로, 융합 잠재표현인 f_3를 출력한다.

[0199] 제4 부호화 블록은 3X3 합성곱 신경망 층 (2↓과 어텐션 모듈(Attention Module)로 구성되며, f3 및 p5 입력에 대한 응답으로, 최종 특징맵 잠재표현 y (f_4)를 출력한다. 한편 어텐션 모듈의 구체적 구성은, 다음 참고 문헌을 참조하여 이해될 수 있다.

[0200] Z. Cheng et. al., "Learned image compression with discretized gaussian mixture likelihoods and attention modules," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7939-7948.

[0201] 입력 특징맵에 패딩을 적용한 것을, 부호화 블록의 입력으로 사용할 수도 있다.

[0202] 구체적으로, 도 17에 도시된 부호화 블록들은 입력되는 소정 계층의 특징맵의 해상도를 절반으로 줄이고, 다음 블록은, 이전 블록의 출력(즉 해상도가 절반으로 감소된 소정 계층의 특징맵)에 소정 계층 대비 해상도가 절반인 다음 계층의 특징맵을 채널 간 결합한 것을 입력으로 사용한다. 이때, 다음 계층의 특징맵의 해상도는 이전

계층의 해상도의 절반이어야 부호화 블록이 동작할 수 있다.

[0203] 구체적으로, 도 17에 도시된 부호화 블록들에 의해, 첫 번째 계층 특징맵 P_2에 대해 최종적으로 4번의 다운샘플링이 수행되므로 P_2의 가로와 세로 해상도는 최소한 16의 배수가 되는 것이 바람직하다. 또한, P_3의 해상도는 8의 배수, P_3의 해상도는 4의 배수, P_5의 해상도는 2의 배수가 되는 것이 바람직할 것이다.

[0204] 위와 같이, 각 계층의 특징맵의 해상도가 특정한 수의 배수가 되어야 하나, 만약, 입력 특징맵의 해상도가 이를 만족하지 못하는 경우, 입력 특징맵의 해상도가 위 제한사항을 만족하도록(즉, 특정한 수의 배수가 되도록), 특징맵에 패딩을 수행하고, 패딩이 수행된 특징맵을 부호화 블록의 입력으로 사용할 수 있다.

[0205] 본 발명의 패딩 과정에 의해 입력 특징맵의 크기를 특정한 수의 배수가 되도록 하는 일 실시예를 자세히 설명한다.

[0206] 입력 특징맵의 가로 길이를 w, 세로 길이를 h라 하고, 변수 m은, 2의 지수(2^n)으로 표현되는 자연수인 것으로 가정한다. w 및 h가 m이 배수가 아닌 경우, 패딩을 통해, 수정된 w' 및 h'가 m의 배수가 되도록 한다.

[0207] 구체적으로, 패딩을 통해 특징맵의 가로 길이에 더해져야 할 픽셀 수를 p_w라 하고, 패딩을 통해 특징맵의 세로 길이에 더해져야 할 픽셀 수를 p_h라 했을 때, p_w 및 p_h는, 다음의 수학식 7 및 8에 의해 계산될 수 있다. 수학식 7 및 8에서, %는, 모듈로 연산을 나타낸다.

수학식 7

[0208] $$p_w = m - (w \% m)$$

수학식 8

[0209] $$p_h = m - (h \% m)$$

[0210] 수학식 7을 통해 계산된 p_w (즉, 패딩을 통해 특징맵의 가로 길이에 더해져야 할 픽셀 수)는, 수학식 9의 예에서와 같이, 특징맵의 왼쪽에 패딩되어야 할 픽셀 수 p_w^{left} 및 특징맵의 오른쪽에 패딩되어야 할 픽셀 수 p_w^{right}는의 합을 나타낼 수 있다. 또한, p_w^{left} 및 p_w^{right}는 각각 수학식 10 및 11에 의해 유도될 수 있다. 수학식 10에서, $\lfloor\rfloor$은 내림 연산을 의미하고, //는 정수 나눗셈 (즉, 나눗셈 연산 후 나머지를 버리고 몫만 취함) 연산을 의미할 수 있다.

수학식 9

[0211] $$p_w^{left} + p_w^{right} = p_w$$

수학식 10

[0212] $$p_w^{left} = p_w // 2 = \lfloor p_w \div 2 \rfloor$$

수학식 11

[0213] $$p_w^{right}=p_w-p_w^{left}$$

[0214] 마찬가지로, 수학식 7을 통해 계산된 p_h(즉, 패딩을 통해 특징맵의 세로 길이에 더해져야 할 픽셀 수)는, 특징맵의 상단에 패딩되어야 할 픽셀수 p_h^{up} 및 특징맵의 하단에 패딩되어야 할 픽셀수 p_h^{down} 의 합을 나타낼 수 있다. 또한, p_h^{up} 및 p_h^{down} 은 각각 수학식 12 및 13에 의해 유도될 수 있다.

수학식 12

[0215] $$p_h^{up}=p_h//2=\lfloor p_h \div 2 \rfloor$$

수학식 13

[0216] $$p_h^{down}=p_h-p_h^{up}$$

[0217] 입력 특징맵이 다계층 특징맵일 경우, 특징맵 잠재표현 추출부(112) 또는 제1 내부 코덱의 구조에 따라, m 적절히 설정할 필요가 있다. 구체적으로, 특징맵 잠재표현 추출부(112)가 도 17에 도시된 부호화 블록들을 포함하는 구조를 가질 때, (즉, l번째 계층 특징맵의 해상도가 (l-1)번째 계층 특징맵의 해상도의 절반이 되는 것이 적절한 구조일 경우), 다음의 수학식 14를 만족하도록, 변수 m을 설정할 수 있다. 수학식 14에서, l 및 (l-1)은, 계층의 인덱스를 나타낸다. 즉, m_l은, l번째 번째 계층에서 사용되는 m의 값을 의미한다. 한편, l은, 1부터 L (L은, 계층의 총 개수)까지의 값을 가질 수 있다.

수학식 14

[0218] $$m_l=m_{l-1}\div 2$$

[0219] 추가적으로, 특징맵 잠재표현 추출부(112) 및 특징맵 복원부(222)를 학습할 때 사용되는 제1 내부 코덱의 종류에 따라, 특징맵 잠재표현의 해상도에 제약이 있을 수 있다. 구체적으로, 제1 내부 코덱에서 두 번의 다운샘플링이 적용되는 경우, 추출된 특징맵 잠재표현 (즉, 도 17에서 y)은 최소 4 이상의 가로와 세로 해상도를 가져야 한다. 따라서, 입력 특징맵 중 해상도가 가장 작은 마지막 계층(L계층)의 특징맵의 가로와 세로 해상도는 최소 8이 되어야 한다. 즉, m_L은, 8과 같거나 큰 값을 갖는 것이 바람직하다.

[0220] 산출된 패딩 크기를 기반으로, 입력 특징맵에 패딩을 수행할 수 있다. 구체적으로, 특징맵의 왼쪽에 p_w^{left}, 오른쪽에 p_w^{right}, 위쪽에 p_h^{up} , 아래쪽에 p_h^{down} 만큼의 픽셀을 추가할 수 있다.

[0221] 한편, 패딩 영역에 추가되는 픽셀의 값은, 기 정의된 값을 가질 수 있다. 여기서, 기 정의된 값은, 0일 수 있다. 또는, 특징 맵의 가장자리에 있는 픽셀 값을 복사 또는 반사한 것을, 패딩 영역에 추가될 픽셀의 값으로 설정할 수도 있다.

[0222] 여기서, 반사는, 패딩 영역 내 n 번째 라인을, 특징맵 내 경계로부터 n 번째 라인에 속하는 픽셀값들로 채우는 것을 의미한다. 여기서, 특징맵 내 경계는, 패딩 영역에 인접하는 경계를 나타낼 수 있다.

[0223] 일 예로, 좌측 패딩 영역의 n 번째 라인은, 특징맵의 좌측 경계로부터, n번째 라인에 속하는 픽셀값들로 채워질 수 있고, 우측 패딩 영역의 n번째 라인은, 특징맵의 우측 경계로부터, n번째 라인에 속하는 픽셀값들로 채워질

수 있다.

[0224] 한편, 특징맵 복원부(222, 224)에서는, 복원된 특징맵에 대해, 언패딩 과정을 수행할 수 있다. 즉, 복원된 특징맵에서, 패딩된 영역을 제거함으로써, 원본 특징맵과 동일한 해상도인 특징맵을 복원할 수 있다. 구체적으로, 복원된 특징맵의 왼쪽 p_w^{left} 크기의 영역, 오른쪽 p_w^{right} 크기의 영역, 위쪽 p_h^{up} 크기의 영역, 아래쪽 p_h^{down} 크기의 영역을 제거함으로써, 원본 특징맵과 동일한 크기의 특징맵을 획득할 수 있다.

[0225] 도 17에 도시된 부호화 블록뿐만 아니라, 영상 처리를 위한 신경망 기반 부호화 블록들을 조합하여, 특징맵 잠재 표현 추출부를 구성할 수도 있다.

[0226] **[D1] 특징맵 잠재 표현을 특징맵으로 복원하는 단계**

[0227] 미세 조정 가능한 특징맵 복원부(224)는, 포맷 역변환된 특징맵 잠재 표현을 원래의 특징맵으로 복원한다. 구체적으로, 미세 조정 가능한 특징맵 복원부(224)는, 학습된 신경 파라미터를 이용하여, 포맷 역변환된 특징맵 잠재 표현을, 원래의 특징맵으로 복원할 수 있다.

[0228] 복원된 결과는, 기계 작업의 수행을 위해 사용되거나, 특징맵 복원부(224)의 학습을 위해 사용될 수 있다.

[0229] 기계 작업의 수행을 위한 특징맵은, 영상 부호화부(140) 및 영상 복호화부(150)에 의해 부호화/복호화되는 반면, 특징맵 복원부(224)의 학습을 위한 특징맵은, 학습 가능한 엔트로피 부호화부(150) 및 학습 가능한 엔트로피 복호화부(250)에 의해 부호화/복호화될 수 있다.

[0230] 한편, 특징맵 복원부(224)의 학습은, 압축 파라미터 무관 특징맵 복원부(228) 및 압축 파라미터 유관 특징맵 복원부(226) 각각에 대해 수행될 수 있다. 이하, 특징맵 복원부(224)의 학습 과정에 대해 상세히 설명하기로 한다.

[0231] 도 18 내지 도 23은 본 개시에 따른 압축 파라미터 무관 특징맵 복원부 및 압축 파라미터 유관 특징맵 복원부를 포함하는 미세 조정이 가능한 특징맵 복원부의 구성도를 나타낸 것이다.

[0232] 도 18에 도시된 예는, 특징맵 복원부(224)의 입력부 쪽의 역이득 유닛 및 어텐션 모듈을 압축 파라미터 유관 특징맵 복원부(226), 구체적으로는, 압축 파라미터 적응부(226-2)로서 구성하는 예를 나타낸 것이다.

[0233] 또는, 도 19에 도시된 예에서와 같이, 압축 파라미터 유관 특징맵 복원부(226)는, 도 18의 예에, 채널 적응부(226-1)를 더 포함하도록 구성될 수도 있다.

[0234] 다른 예로, 도 20에 도시된 예에서와 같이, 역이득 유닛만을 압축 파라미터 유관 특징맵 복원부(226)의 구성으로 이용하고, 어텐션 모듈은 압축 파라미터 무관 특징맵 복원부(228)의 구성으로 포함시킬수도 있다.

[0235] 또는, 도 21에 도시된 예에서와 같이, 압축 파라미터 유관 특징맵 복원부(226)는, 도 20의 예에, 채널 적응부(226-1)를 더 포함하도록 구성될 수도 있다.

[0236] 다른 예로, 도 22에 도시된 예에서와 같이, 채널 적응부(226-1)만을 압축 파라미터 유관 특징맵 복원부(226)의 구성으로 설정할 수 있다. 또한, 일반적인 특징맵 복원부(222)를 압축 파라미터 무관 특징맵 복원부(228)로 설정할 수 있다.

[0237] 한편, 미세 조정 가능한 특징맵 복원부(224)는, 압축 파라미터 유관 특징맵 복원부(226) 없이 압축 파라미터 무관 특징맵 복원부(228) 만으로 구성될 수도 있다. 도 23은, 밋 조정 가능한 특징맵 복원부(224)가 압축 파라미터 무관 특징맵 복원부(228) 만으로 구성되는 예를 나타낸 것이다.

[0238] 포맷 역변환부(240)에서 출력되는 특징맵 잠재 표현 뿐만 아니라, 영상 복호화부의 압축 파라미터를 추가적으로 미세 조정 가능한 특징맵 복원부(224)로 입력할 수 있다. 즉, 특징맵 잠재 표현에 압축 파라미터를 추가(또는, 결합)한 것을, 미세 조정 가능한 특징맵 복원부(224)에 입력할 수 있다.

[0239] 한편, 압축 파라미터를 채널 간 결합 방식으로 복원된 특징맵 잠재 표현에 추가하기 위해서는, 압축 파라미터를, 복원된 특징맵 잠재표현과 같은 공간 해상도를 갖는 표현으로 변환할 필요가 있다. 이를 위해, 원-핫 인코딩 방식을 적용할 수 있다.

[0240] 도 24는는 원-핫 인코딩 방식으로 압축 파라미터를 변환한 실시예를 나타낸다.

[0241] 원-핫 인코딩 방식이란, 압축 파라메터 값에 해당하는 채널은 모두 1의 값으로 채우고 그렇지 않은 채널은 모두

0의 값으로 채우는 방식을 나타낸다.

[0242] 복원된 특징맵 잠재 표현의 채널 수가 C이고, 변환된 압축 파라미터의 채널 수를 C_{QP}라 할 경우, 복원된 특징맵 잠재 표현 및 변환된 압축 파라미터의 채널 결합에 의해 채널 개수가 C' (C+C_{QP})인 결합 데이터가 생성된다.

[0243] 이는, 미세 조정 가능한 특징맵 복원부(224)에 입력되는 채널들의 개수를 C개에서 C'개로 증가시키는 결과를 초래한다.

[0244] 위와 같은 문제를 해소하기 위해, 채널 개수가 C'인 결합 데이터를 다시 C개의 채널 수로 변환하는 채널 적응부(226-1)를 이용할 수 있다. 즉, 채널 적응부(226-1)는, C'의 채널 수를 갖는 입력 데이터를, C의 채널 수를 갖는 데이터로 변환하기 위한 신경망 기반의 모듈을 의미할 수 있다.

[0245] 한편, 채널 적응부(226-1)를 통과하더라도, 입력되는 데이터의 공간적 해상도는 변경되지 않을 수 있다.

[0246] 도 25는 본 개시의 일 실시예에 따른 채널 적응부를 나타낸 것이다.

[0247] 한편, 미세 조정 가능한 특징맵 복원부(224)는, 영상 복호화부(230)에서 이용 가능한 압축 파라미터들에 대해 학습을 수행할 수 있다.

[0248] 도 26 및 도 27은 미세 조정 가능한 특징맵 복원부가 영상 복호화부에서 이용 가능한 압축 파라미터들에 대해 미세 조정 학습되는 예를 나타낸 것이다.

[0249] 도 26은 미세 조정 가능한 특징맵 복원부(224)가 영상 복호화부(230)에서 이용 가능한 모든 압축파라미터들에 대해 미세 조정 학습되는 예를 나타낸 것이다.

[0250] 영상 복호화부에서 이용 가능한 모든 압축 파라미터들에 대한 학습을 수행함으로써, 여러 압축 파라미터들에 대응 가능한, 단일의 미세 조정 가능한 특징맵 복원부(224)가 획득될 수 있다.

[0251] 이때, 압축 파라미터 유관 특징맵 복원부(226)만이, 영상 복호화부(230)에서 사용 가능한 압축 파라미터들에 기반하여 학습되고, 압축 파라미터 무관 특징맵 복원부(228)는, 압축 파라미터에 관계없이 (즉, 모든 압축 파라미터에 대해), 동일한 신경망 가중치 파라미터를 갖도록 학습될 수 있다.

[0252] 도 27은 영상 복호화부에서 이용 가능한 압축 파라미터들을 그룹화하고, 미세 조정 가능한 특징맵 복원부가 각 그룹을 대상으로 학습되는 예를 나타낸 것이다.

[0253] 그룹 단위로 학습을 수행함으로써, 복수의 그룹들 각각에 대해, 특징맵 복원부(220)의 훈련된 신경망 파라미터가 획득될 수 있다.

[0254] 즉, 실제 영상 부호화부/복호화부(140, 230)에서 사용하는 압축 파라미터가 결정되면, 특징맵 복원부(220)는, 상기 압축 파라미터가 속하는 그룹에 대응되는 신경망 파라미터를 이용하여, 특징맵을 복원할 수 있다.

[0255] 한편, 도 27에서는, 2개의 압축 파라미터들이 하나의 그룹으로 설정되는 것으로 예시되었다.

[0256] 그룹에는 최소 하나 이상의 압축 파라미터를 포함할 수 있고, 또한, 그룹별로, 포함하는 압축 파라미터들의 개수가 서로 상이할 수도 있다.

[0257] 특징맵 복원부(220)는, 기본으로 설정된 특징맵 복원부의 신경망 파라미터를 사용할 수 있다. 또는, 특징맵 복원부(220)에 대해, 특징맵이 추출된 기계 작업 모델의 종류, 기계 작업 종류, 또는 특징맵이 추출된 기계 작업 모델의 분할 지점 중 적어도 하나에 기초하여, 복수의 신경망 파라미터 후보들이 존재할 수 있다. 특징맵 복원부(220)는, 복수의 신경망 파라미터 후보들 중 하나를 선택하여 사용할 수 있다. 이를 위해, 복수의 신경망 파라미터 후보들 중 하나를 지시하는 (예컨대, feat_restoration_weight_idx) 정보가 부호화되어 시그날링 될 수 있다.

[0258] 추가적으로, 특징맵 부호화부(110)에서 패딩(padding)이 수행된 경우, 특징맵 복호화부(220)에서는 언패딩(unpadding)이 수행되는 것이 바람직하다. 언패딩 과정을 위해, 패딩이 수행되기 전의 입력 특징맵의 해상도(예컨대, ori_feat_wid, ori_feat_hei)가 부호화되어 시그날링될 수 있다.

[0259] 패딩이 수행된 후의 특징맵의 가로 길이를 w_{pad}, 패딩이 수행된 후의 특징맵의 세로 길이를 h_{pad} 라 가정하자. 이때, w_{pad} 및 h_{pad}는, 복원된 잠재표현 특징맵의 해상도, 계층 인덱스, 특징맵 복원부의 구성을 통해 계산될 수 있다. 예를 들어서, 특징맵 복원부(224)가, 도 23에 도시된 예와 같이 구성되고, 복원된 잠재표현 특징맵의 가

로와 세로 해상도가 4인 것으로 가정한다. 다계층 특징맵 내 첫 번째 계층의 특징맵(즉, $\hat{p}_2$)를 복원하기 위해서는, 4번의 업샘플링 (즉, 가로와 세로 해상도를 16배씩 증가시킴)을 적용하기 때문에 $\hat{p}_2$의 w_{pad} 및 h_{pad}의 값은 64로 결정될 수 있다.

[0260] 또한, 패딩이 수행되기 전의 특징맵의 가로와 세로 길이를 각각 w 및 h 라고 하면 수학식 15과 수학식 16을 통해 패딩을 통해 특징맵의 가로에 더해진 픽셀 수 p_w 및 패딩을 통해 특징맵의 세로에 더해진 픽셀 수 p_h를 계산할 수 있다.

수학식 15

[0261] $$p_w = w_{pad} - w$$

수학식 16

[0262] $$p_h = h_{pad} - h$$

[0263] 다른 예로, 가로 방향 패딩 크기 p_w 및 세로 방향 패딩 크기 p_h를 유도하기 위해, 패딩 과정에서 사용된 변수 m을 부호화하여 시그날링할 수도 있다. 변수 m이 복호화되면, 수학식 7 및 수학식 8을 통해, 계산할 수 있다.

[0264] 상술한 바에 따라, p_w 및 p_h가 계산되면, 10 내지 13을 통해, 특징맵의 왼쪽, 오른쪽, 위쪽, 아래쪽에 더해진 픽셀 수인 p_w^{left}, p_w^{right}, p_h^{up}, p_h^{down}을 계산할 수 있다. 이후, 패딩된 픽셀 수 만큼을 복원된 특징맵에서 제거함으로써 언패딩 된 특징맵을 구할 수 있다. 구체적으로, 특징맵의 왼쪽에 p_w^{left}, 오른쪽에 p_w^{right}, 위쪽에 p_h^{up}, 아래쪽에 p_h^{down} 만큼의 픽셀을 제거하여 언패딩 과정을 수행할 수 있다.

[0265] **[E2] 포맷 변환 단계 / [D2] 포맷 역변환 단계**

[0266] 포맷 변환부(130)에서는, 추출된 특징맵 잠재 표현을, 영상 부호화부(140)의 입력 포맷에 맞춰 변환할 수 있다.

[0267] 구체적으로, 포맷 변환 과정은, 특징맵 잠재 표현의 채널들 배열하는 단계 및 특징맵 잠재 표현을 양자화하는 단계를 포함할 수 있다.

[0268] 포맷 역변환부(240)에서는, 영상 복호화부(230)에서의 출력을, 특징맵 부호화부(110)로부터터 추출된 특징맵 잠재표현과 동일한 포맷으로 변환한다.

[0269] 구체적으로, 포맷 역변환 과정은, 포맷 변환된 잠재 표현을 역양자화하는 단계 및 포맷 역변환된 잠재 표현(즉, 복원된 잠재 표현)의 채널들을 재배열하는 단계를 포함할 수 있다.

[0270] 한편, 영상 부호화기에서 수행된 포맷 변환의 역변환을 수행하기 위해, 포맷 변환 정보가 영상 복호화기로 전달되어야 한다.

[0271] 일 예로, 특징맵 잠재 표현을 양자화하기 위해 사용되는 비트수 n 및 잠재 표현의 범위 $Range_{max}$가 부호화되어, 포맷 역변환부(240)의 역양자화부에 전달될 수 있다.

[0272] 또한, 특징맵 잠재 표현의 채널들의 배열 방법을 지시하는 정보가 부호화되어 시그날링될 수 있다. 배열 방법은, 공간적 배열 방법, 시간적 배열 방법 또는 시간공간 배열 방법 중 하나를 지시할 수 있다.

[0273] 한편, 공간적 배열 방법 또는 시공간적 배열 방법이 이용된 경우에는, 하나의 프레임을 구성하는 가로 채널 개수 및 세로 채널 개수를 나타내는 정보가 추가 부호화되어 시그날링될 수 있다.

[0274] 또는, 위와 같은, 포맷 변환 정보들이 부호화기 및 복호화기에서 기 정의되어 있을 수도 있다.

[0275] 이하, 포맷 변환 및 역변환 방법에 대해 상세히 살펴보기로 한다.

[0276] [E2-1 / D2-2] 특징맵 잠재 표현의 채널들을 배열 / 재배열하는 방법

[0277] 도 28은 일반적인 특징맵의 구성을 예시한 것이다.

[0278] 도 28에 도시된 예에서와 같이, 특징맵은 복수개의 채널들로 구성되어 있을 수 있다. 구체적으로, 도 28에서는, 채널들의 개수를 N_c, 채널의 가로 크기(즉, 너비)를 W_c, 채널의 세로 크기(즉, 높이)를 H_c로 표시하였다.

[0279] 복수개의 채널들을 부호화하기 위해, 복수개의 채널들을, 영상 부호화부의 입력 단위인 프레임(즉, 픽처) 형태로 변환할 필요가 있다.

[0280] 이를 위해, 공간적 배열 방법, 시간적 배열 방법 또는 시공간적 배열 방법 중 적어도 하나를 이용할 수 있다.

[0281] 도 29 내지 도 31은, 각각 공간적 배열 방법, 시간적 배열 방법 및 시공간적 배열 방법을 나타낸 것이다.

[0282] 도 29에 도시된 예에서와 같이, 공간적 배열 방법은, 특징맵 잠재 표현의 채널들을 마치 타일과 형태로 배열함으로써, 하나의 특징맵 잠재 표현 프레임을 구성하는 것을 나타낸다.

[0283] 가로 방향의 배열 개수 m 및 세로 방향의 배열 개수 n은, 상호간의 곱이, 배열한 채널 개수 Nc와 같아지도록 설정될 수 있다.

[0284] 가로 방향의 배열 개수 m 및 세로 방향의 배열 개수 n을 결정하기 위한 정보가 부호화되어 시그날링될 수 있다.

[0285] 또는, 영상 복호화기(20)에는, 비트스트림으로부터 복호화된 영상의 크기 정보(예컨대, pic_width_in_luma_samples 및 pic_height_in_luma_samples)를 복호화하고, 특징맵의 해상도를 기반으로, 픽처 내 채널들의 배열, 즉, 가로 방향 배열 개수 및 세로 방향 배열 개수를 획득할 수 있다. 일 예로, 픽처의 너비(즉, pic_width_in_luma_samples)를 특징맵의 너비로 나눈 값이 가로 방향 배열 개수 m으로 유도되고, 픽처의 높이(즉, pic_height_in_luma_samples)를 특징맵의 높이로 나눈 값이 세로 방향 배열 개수 n으로 유도될 수 있다.

[0286] 공간적 배열된 특징맵 잠재 표현 프레임은, 영상 부호화부 및 영상 복호화부에서, 화면 내 예측(Intra-frame prediction) 방법에 기반하여 부호화/복호화될 수 있다.

[0287] 도 30에 도시된 예에서와 같이, 시간적 배열 방법은, 특징맵 잠재 표현의 채널들 각각이 하나의 프레임이 되도록 나열하는 것을 나타낸다.

[0288] 이에 따라, 하나의 특징맵 잠재 표현을 시간적으로 배열함으로써 생성되는 프레임들의 총 개수는, 특징맵 잠재 표현을 구성하는 채널들의 총 개수 N_c와 같을 수 있다.

[0289] 시간적 배열된 특징맵 잠재 표현 프레임은, 영상 부호화부 및 영상 복호화부에서, 화면 간 예측(Inter-frame prediction) 방법에 기반하여 부호화/복호화될 수 있다. 한편, 시간적 배열은, 채널들을 시간적으로 배열하여 생성된 것인바, 화면 내 예측을, 채널 간 예측(Intra-Channel prediction)이라 호칭할 수도 있다.

[0290] 도 31에 도시된 예에서와 같이, 시공간적 배열 방법은, 공간적으로 배열된 프레임들을 시간적으로 배열한 것을 나타낸다.

[0291] 하나의 프레임을 구성하는 채널들의 개수 mxn 및 총 프레임의 수를 곱한 값이 N_c와 같아지도록 설정될 수 있다.

[0292] 변환된 특징맵 프레임은, 영상 부호화부(130)에 의해 압축되고, 영상 복호화부(230)에 의해 복원된다.

[0293] 영상 복호화부(230)에서 복호화된 특징맵 잠재 표현 프레임은, 특징맵 재배열을 통해, 원래의 채널 구성 형태로 복원될 수 있다.

[0294] [E2-2 / D2-1] 변환된 특징맵 잠재 표현 프레임을 양자화 / 역양자화하는 방법

[0295] 특징맵 잠재 표현은, 균등 양자화 또는 비균등 양자화 방식으로 양자화될 수 있다. 구체적으로, 특징맵 잠재 표현은, n-bit의 정수로 양자화될 수 있다.

[0296] 수학식 17은, 특징값 $F_{original}$의 n-bit 균등 양자화 과정을 나타낸다.

수학식 17

[0297]
$$F_{converted}=Round((F_{original}\times(2^{n}-1)/Range_{\max}+2^{n-1}$$
$$If,F_{converted}>2^{n-1},F_{converted}=2^{n-1},$$
$$F_{converted}<0,F_{converted}=0$$

[0298] 수학식 18은, 복호화된 특징값 $F_{converted}$의 균등 역양자화 과정을 나타낸다.

수학식 18

[0299]
$$\hat{F}=(\frac{(\tilde{F}_{Converted}-2^{n-1})}{(2^{n}-1)})\times Range_{max}$$

[0300] 양자화시, 정수화 과정에서 발생하는 오차를 줄이기 위해, 반올림(Rounding) 연산이 적용될 수 있다. 또한, n-bit 정수로 표현 가능한 범위를 넘는 값들을 클리핑(Clipping)하는 과정이 추가될 수도 있다.

[0301] 수학식 17 및 수학식 18에 예시된 바와 같이, 양자화 및 역양자화를 수행하기 위해서는, 비트수 n 및 특징값의 범위 $Range_{max}$ 값을 요구된다. 영상 복호화기에서 영상 부호화기에서 동일한 비트수 n 및 특징값의 범위 $Range_{max}$를 사용하기 위해, 비트수 n 및 특징값의 범위 $Range_{max}$ 가 영상 부호화기 및 영상 복호화기에서 기 정의되어 있을 수 있다. 또는, 비트수 n 및 특징값의 범위 $Range_{max}$ 중 적어도 하나를 유도하기 위한 정보가 부호화되어 영상 복호화기로 시그날링될 수 있다.

[0302] 비트수 n은 8 또는 10과 같이, 영상 부호화기에 입력 가능한 값으로 설정할 수 있다. 또는, 압축 성능 향상을 위해, 4 또는 6과 같이, 더 작은 값으로 설정할 수 있다.

[0303] $Range_{max}$ 값은, 부호화할 특징값들 중 최대값에서 최소값을 차분하여 유도할 수 있다. 또는, 기 정의된 값을 $Range_{max}$ 값으로 설정할 수 있다.

[0304] 또는, 비트수 n 및 특징값들 중 최대값 및 최소값을 기반으로, 선형 양자화를 수행할 수도 있다.

[0305] 수학식 19는, 복호화된 특징값 $F_{converted}$에 대해 역양자화를 수행하는 예를 나타낸다.

수학식 19

[0306]
$$DequantizedPackedFeature=F_{converted}/(2^{(n-1)})*(F_val_max-F_val_min)+F_val_min$$

[0307] 수학식 19에서, F_val_max는, 특징값들 중 최대값을 나타내고, F_val_min은, 특징값들 중 최소값을 나타낸다. 역양자화를 수행하기 위해, 특징값들 중 최대값 및 최소값을 나타내는 정보가 부호화되어 시그날링될 수 있다.

[0308] 한편, 양자화는, 잠재 표현 배열 이후에 수행될 수도 있고, 잠재 표현 배열 이전에 수행될 수도 있다. 마찬가지로, 잠재 표현의 역양자화도, 잠재 표현의 재배열 이전에 수행될 수도 있고, 잠재 표현의 재배열 이후에 수행될 수도 있다.

[0309] 한편, 잠재 표현 배열 이전에 양자화가 수행되었다면, 잠재 표현 재배열 이전에 역양자화를 수행하여, 계산 오차를 줄이는 것이 바람직할 것이다.

[0310] **[E3]/[D3] 영상 부호화 단계 및 영상 복호화 단계**

[0311] 영상 부호화기(10) 내 영상 부호화부(140, 150)에서는, 변환된 특징맵 잠재 표현을 비트스트림으로 변환한다.

[0312] 영상 복호화부(230, 250)는, 비트스트림을 변환된 특징맵 잠재 표현으로 복원할 수 있다.

[0313] 일 예로, 기계 작업 수행을 위한 영상 부호화/복호화시에는, 영상 부호화부(140) 및 영상 복호화부(230)가 이용되고, 부호화 모델을 학습하기 위한 영상 부호화/복호화시에서는, 학습 가능한 영상 부호화부(150) 및 학습 가능한 영상 복호화부(250)가 이용될 수 있다.

[0314] 한편, 영상 부호화부(140) 및 영상 복호화부(230)에서는, HEVC 또는 VVC와 같은 영상 압축 코덱을 사용할 수 있다.

[0315] 본 개시에서 설명하는 영상 부호화 / 복호화 방법을 지원하기 위해, 다음과 같은, 신택스 구문 / 신택스 요소들이 정의될 수 있다.

[0316] 표 1은, 입력 영상 및 스케일된 영상의 크기 정보를 포함하는, 시각 모델(Vision model) 파라미터 세트를 예시한 것이다.

표 1

[0317]

vision_model_parameter_set_rbsp ( ) {
vmps_vision_model_parameter_set_id
img_wid
img_hei
scaled_img_wid
scaled_img_hei
}

[0318] 표 1에서, 신택스 vmps_vision_model_parameter_set_id는, VMPS 구조에 할당되는 식별자를 나타낸다.

[0319] 신택스 img_width은, 입력 영상의 너비를 나타내고, img_hei는 입력 영상의 높이를 나타낸다.

[0320] 신택스 scaled_img_wid는, 스케일된 입력 이미지의 너비를 나타내고, scaled_img_hei는, 스케일된 입력 이미지의 높이를 나타낸다.

[0321] 표 2는, 피처 시퀀스 파라미터 세트를 예시한 것이다.

표 2

[0322]

feat_seq_parameter_set_rbsp ( ) {
fsps_feat_seq_parameter_set_id
fsps_vision_model_parameter_set_id
num_ori_feat_layers
for (i = 0; i <= num_feature_layers; i++) {
ori_feat_wid[i]
ori_feat_hei[i]
num_ori_feat_chan[i]
}
fused_feat_wid
fused_feat_hei
}

[0323] 표 2에서, 신택스 fsps_feat_seq_parameter_set_id는, FSPS (Feature Sequence Parameter Set)에 할당되는 식별자를 나타낸다.

[0324] 신택스 fsps_vision_model_parameter_set_id는, FSPS가 참조하는 비전 모델 파라미터 세트의 식별자를 나타낸다.

[0325] 신택스 num_ori_feature_layers는, 복원될 특징 계층들의 개수를 나타낸다. 즉, 신택스 num_ori_feature_layers는, 다계층 특징맵의 계층들의 개수를 나타낸다.

[0326] 신택스 ori_feat_wid[i]는 인덱스가 i인 계층 내 원본 특징맵의 너비를 나타낸다.

[0327] 신택스 ori_feat_hei[i]는 인덱스가 i인 계층 내 원본 특징맵의 높이를 나타낸다.

[0328] 신택스 num_ori_feat_chan[i]는 인덱스가 i인 계층 내 원본 특징맵의 채널 개수를 나타낸다.

[0329] 한편, i는 0부터, (num_ori_feat_wid - 1) 까지의 값을 가질 수 있다.

[0330] 신택스 fused_feat_wid는, 융합된 특징맵의 너비를 나타낸다. 여기서, 융합된 특징은, 특징맵 부호화부(110)를 통해 추출된 특징맵 잠재표현을 의미한다.

[0331] 신택스 fused_feat_hei는, 융합된 특징맵의 높이를 나타낸다.

[0332] 표 3은, 피처 픽처 파라미터 세트를 예시한 것이다.

표 3

[0333]

feature_pic_parameter_set_rbsp() {
fpps_feat_pic_parameter_set_id
fpps_feat_seq_parameter_set_id
inner_decoding_bypass_flag
if(!inner_decoding_bypass_flag){
feat_inner_decoder_info()
}
dequant_bypass_flag
if(!dequant_bypass_flag) {
packed_feat_val_max
packed_feat_val_min
}
unpacking_bypass_flag
if(!unpacking_bypass_flag) {
symmetric_feat_chan_flip_bypass_flag
}
feat_restoration_bypass_flag
if(!feat_restoration_bypass_flag)
feat_restoration_info()
}
}

[0334] 표 3에서, 신택스 fpps_feat_pic_parameter_set_id는, FPPS (Feature Picture Parameter Set)에 할당되는 신택스를 나타낸다.

[0335] 신택스 fpps_feat_seq_parameter_set_id는, FPPS가 참조하는 FSPS의 식별자를 나타낸다.

[0336] 신택스 inner_coding_bypass_flag는 이너 코덱이 동작하는지 여부를 나타낸다. 일 예로, 신택스 inner_coding_bypass_flag가 1인 경우, 특징맵 잠재 표현 추출부(112)에서 추출된 특징맵 잠재 표현이 그대로 특징맵 복원부(222)로 전송될 수 있다.

[0337] 반면, 신택스 inner_coding_bypass_flag가 0인 경우, 특징맵 잠재 표현 추출부(112)에서 추출된 특징맵 잠재 표현이, 내부 부호화부를 통해 부호화되고, 부호화 결과 생성된 비트스트림이 영상 복호화기(20)로 전송될 수 있다.

[0338] 신택스 dequant_bypass_flag는, 양자화/역양자화가 수행되었는지 여부를 나타낸다. 신택스 dequant_bypas_flag가 1인 것은, 양자화/역양자화의 수행이 생략됨을 나타낸다. 반면, 신택스 dequant_bypass_flag가 0인 것은, 양자화/역양자화가 수행됨을 나타낸다.

[0339] 신택스 dequant_bypass_flag가 0인 경우, 양자화/역양자화 수행과 관련된 정보가 추가 부호화/복호화될 수 있다.

[0340] 일 예로, 신택스 packed_feat_val_max는, 배열된 프레임 내 특징값들 중 최대값을 나타낸다. 또한, 신택스 packed_feat_val_min는, 특징값들 중 최소값을 나타낸다.

[0341] 신택스 unpacking_bypass_flag는, 특징 재배열이 수행되어야 하는지 여부를 나타낸다. 신택스 unpacking_bypass_flag가 1인 것은, 특징 배열 / 재배열이 수행되지 않음을 나타낸다. 반면, 신택스 unpacking_bypass_flag가 0인 것은, 특징 배열/재배열이 수행됨을 나타낸다.

[0342] 신택스 unpacking_bypass_flag가 0인 경우, 포맷 역변환부(240)에서 특징 재배열을 수행하기 위한 정보가 추가 부호화/복호화될 수 있다.

[0343] 일 예로, 신택스 symmetric_feat_chan_flip_bypass_flag는, 대칭형 특징 채널 플리핑(flipping) 재배열이 허용되는지 여부를 나타낸다.

[0344] 신택스 feature_restoration_bypass_flag는, 특징맵 부호화부 및 특징맵 복호화부의 수행 여부를 나타낸다. 일 예로, 신택스 feature_restoration_bypass_flag가 1인 것은, 특징맵 부호화부 및 복호화부가 수행되지않음을 나타낸다.

[0345] 반면, 신택스 feature_restoration_bypass_flag가 0인 것은, 특징맵 부호화부 및 특징맵 복호화부가 수행됨을나타낸다. 신택스 feature_restoration_bypass_flag가 0인 경우, 특징맵 복호화부의 신경망 파라미터를 결정하기 위한정보가 추가 부호화/복호화될 수 있다.

[0346] 표 4는 내부 코덱과 관련된 신택스를 포함하는 구문 구조를 예시한 것이다.

표 4

[0347]

feat_inner_decoder_info( ) {
inner_coding_idx

[0348] 표 4에서, 신택스 inner_coding_idx는, 내부 부호화부 및 내부 복호화부에서 이용하는 코덱의 종류를 나타낸다. 신택스 inner_coding_idx는, 복수의 코덱들 중 하나를 지시할 수 있다. 일 예로, 신택스 inner_coding_idx의 값이 0인 것은, 내부 부호화부 및 내부 복호화부에서, 영상을 부호화/복호화하기 위해 VVC가 사용됨을 나타낸다.

[0349] 표 5는 특징 복원과 관련된 신택스를 포함하는 구문 구조를 예시한 것이다.

표 5

[0350]

feat_restoration_info( ) {
feat_restoration_weight_idx
pad_size_min
}

[0351] 표 5에서 신택스 feat_restoration_weight_idx는, 특징맵 복호화부에서 사용될 신경망 파라미터를 특정한다. 구체적으로, feat_restoration_weight_idx의 값이 0인 것은, 탐색 작업 (detection task)을 위한 신경망 파라미터가 사용됨을 나타낸다. feat_restoration_weight_idx의 값이 1인 것은, 분할 작업(segmentation task)을 위한 신경망 파라미터가 사용됨을 나타낸다. feat_restoration_weight_idx의 값이 2인 것은, 객체 추적 작업 (object tracking task) 을 위해, 기계 작업 모델로 YOLO가 사용된 경우, 'DN53(DarkNet53)'의 기계 작업 모델 분할 지점으로부터 추출된 특징 (즉, MPEG FCM의 Common Test and Training Conditions에서 TVD 데이터 세트를 위해 정의된 분할 지점) 을 위한 신경망 파라미터가 사용됨을 나타낸다. feat_restoration_weight_idx의 값이 3인 것은, 객체 추적 작업 (object tracking task) 을 위해, 기계 작업 모델로 YOLO가 사용된 경우, 'ALT1'의 기계 작업 모델 분할 지점으로부터 추출된 특징 (즉, MPEG FCM의 Common Test and Training Conditions에서 HiEve 데이터 세트를 위해 정의된 분할 지점)을 위한 신경망 파라미터가 사용을 나타낸다.

[0352] 신택스 pad_size_min은, 가장 작은 해상도를 갖는 계층 적용되는 패딩의 양(즉, 패딩의 크기)를 나타낸다.

[0353] 상술한 실시예들에서 소개된 신택스 요소들의 명칭은, 본 개시에 따른 실시예들을 설명하기 위해 임시로 부여된 것에 불과하다. 본 개시에서 제안된 것과 상이한 이름으로 신택스 요소들을 명명할 수도 있다.

[0354] 본 개시의 예시적 실시예들에서 설명된 구성요소는, 하드웨어 요소에 의해 구현될 수 있다. 예를 들어, 상기 하드웨어 요소는, 디지털 신호 프로세서(DSP), 프로세서, 컨트롤러, 주문형 집적 회로(ASIC), FPGA와 같은 프로그

래머블 로직 요소, GPU, 기타 전자 장치, 또는 이들의 조합 중 적어도 하나를 포함할 수 있다. 본 개시의 예시적 실시예들에서 설명된 기능 또는 과정 중 적어도 일부는 소프트웨어로 구현될 수 있으며, 소프트웨어는 기록 매체에 기록될 수 있다. 예시적 실시예들에서 설명하는 구성요소, 기능 및 프로세스는 하드웨어와 소프트웨어의 조합으로 구현될 수 있다.

[0355] 본 개시의 일 실시예에 따른 방법은 컴퓨터에 의해 수행될 수 있는 프로그램으로 구현될 수 있으며, 상기 컴퓨터 프로그램은 자기 저장 매체, 광 판독 매체, 디지털 저장 매체 등과 같은 다양한 기록 매체에 기록될 수 있다.

[0356] 본 개시에서 설명한 다양한 기술들은, 디지털 전자 회로 또는 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 또는 이들의 조합으로 구현될 수 있다. 상기 기술들은 컴퓨터 프로그램 제품, 즉, 정보 매체에 유형적으로 구현되는 컴퓨터 프로그램 또는 컴퓨터 프로그램(예를 들어, 기계 판독 가능 저장 장치(예: 컴퓨터 판독 가능 매체) 또는 데이터 처리 장치), 데이터 처리 장치에 의해 처리되거나 데이터 처리 장치(예를 들어, 프로그램 가능한 프로세서, 컴퓨터 또는 복수의 컴퓨터들)의 동작을 하기 위해 전파된 신호로 구현되는 컴퓨터 프로그램으로 구현될 수 있다.

[0357] 컴퓨터 프로그램(들)은 컴파일된 언어나 해석된 언어를 포함한 모든 형태의 프로그래밍 언어로 작성될 수 있으며, 독립 실행형 프로그램 또는 모듈, 구성 요소, 서브루틴 또는 컴퓨팅 환경에서 사용하기에 적합한 기타 단위를 포함하는 모든 형태로 배포될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨터, 또는 한 사이트 또는 여러 사이트에 분산되어 있고 통신 네트워크로 상호 연결되어 있는 복수 컴퓨터들에 의해 수행될 수 있다.

[0358] 컴퓨터 프로그램의 실행에 적합한 프로세서의 예로, 범용 및 특수 목적 마이크로프로세서, 및 디지털 컴퓨터의 하나 이상의 프로세서가 포함된다. 일반적으로 프로세서는 읽기 전용 메모리나 랜덤 액세스 메모리 또는 둘 모두에서 명령과 데이터를 수신한다. 컴퓨터의 구성 요소는 명령을 실행하기 위한 적어도 하나의 프로세서, 및 명령 및 데이터를 저장하기 위한 하나 이상의 메모리 장치를 포함할 수 있다. 또한, 컴퓨터는 데이터 저장을 위한 하나 이상의 대용량 저장 장치, 예를 들어, 자기, 광자기 디스크 또는 광 디스크를 포함하거나, 상기 대용량 저장 장치에 연결되어 데이터를 수신 및/또는 전송을 수행할 수 있다. 컴퓨터 프로그램 명령 및 데이터를 구현하기에 적합한 정보 매체의 예로는, 반도체 메모리 장치(예를 들어, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체), 컴팩트 디스크 읽기 전용 메모리(CD- ROM), 디지털 비디오 디스크(DVD) 등과 같은 광학 매체, 플롭티컬 디스크와 같은 광자기 매체, 및 ROM(Read Only Memory), RAM(Random Access Memory), 플래시 메모리, EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 및 기타 알려진 컴퓨터 판독 가능 매체가 포함된다. 프로세서와 메모리는 특수 목적 논리 회로에 의해 보완되거나 통합될 수 있다.

[0359] 프로세서는 운영 체제(OS) 및 OS에서 실행되는 하나 이상의 소프트웨어 애플리케이션을 실행할 수 있다. 프로세서 장치는 또한 소프트웨어 실행에 응답하여 데이터를/에 액세스, 저장, 조작, 처리 및 생성할 수 있다. 단순화를 위해 프로세서 장치는 단수로 설명되었으나, 해당 기술 분야의 통상의 지식을 가진자는, 프로세서 장치가 복수의 처리 요소들 및/또는 다양한 타입의 처리 요소들을 포함할 수 있음을 이해할 수 있다. 예를 들어, 프로세서 장치는 복수 프로세서들 또는 프로세서(processor)와 제어부(controller)를 포함할 수 있다. 또한, 병렬 프로세서들과 같이 상이한 처리 구조를 구성할 수도 있다. 또한, 컴퓨터가 읽을 수 있는 매체는 컴퓨터가 접근할 수 있는 모든 매체를 의미하며, 컴퓨터 저장매체와 전송매체를 모두 포함할 수 있다.

[0360] 본 개시는 다양한 상세 구현 예들의 세부 설명을 포함하나, 그 세부 사항이 본 개시에서 제안하는 발명 또는 청구 범위를 제한하는 것은 아니고, 특정한 예시적 실시예의 특징을 설명하는 것으로 이해되어야 한다.

[0361] 본 개시에서 예시적 실시예들에서 개별적으로 설명된 특징들은 단일 예시적 실시예에 의해 구현될 수 있다. 반대로, 본 개시에서 단일 예시적 실시예에 대해 설명된 다양한 특징들은 복수의 예시적 실시예들의 조합 또는 적절한 하위 조합에 의해 구현될 수도 있다. 나아가, 본 개시에서, 상기 특징들은 특정 조합에 의해 동작할 수 있고, 최초 상기 조합이 청구된 것처럼 기술될 수 있으나, 경우에 따라, 하나 이상의 특징이 청구된 조합에서 제외되거나, 청구된 조합이 하위 조합 또는 하위 조합의 수정 형태로 변경될 수 있다.

[0362] 마찬가지로, 도면에 특정 순서로 동작이 설명되어 있다 하더라도, 목표된 결과를 획득하기 위해, 동작들을 특정 순번 또는 순서로 실행하는 것이 필요하거나, 모든 동작들의 수행이 필요한 것으로 이해되어서는 안 된다. 특정한 경우에는 멀티태스킹과 병렬 처리가 유용할 수 있다. 또한, 모든 실시예들의 예시적 실시예들에서 다양한 장치 구성요소들이 반드시 분리되어야 하는 것으로 이해되어서는 안 되며, 전술한 프로그램 구성요소 및 장치는

단일 소프트웨어 제품 또는 여러 소프트웨어 제품에 패키징될 수 있다.

[0363] 본 명세서에 개시된 예시적 실시예들은 단지 예시적인 것이며 본 개시의 범위를 제한하려는 것이 아니다. 해당 기술 분야의 통상의 지식을 가진자는, 청구범위 및 그 균등물의 사상 및 범위를 벗어나지 않으면서 예시적 실시예들에 대한 다양한 수정이 이루어질 수 있음을 인식할 수 있을 것이다.

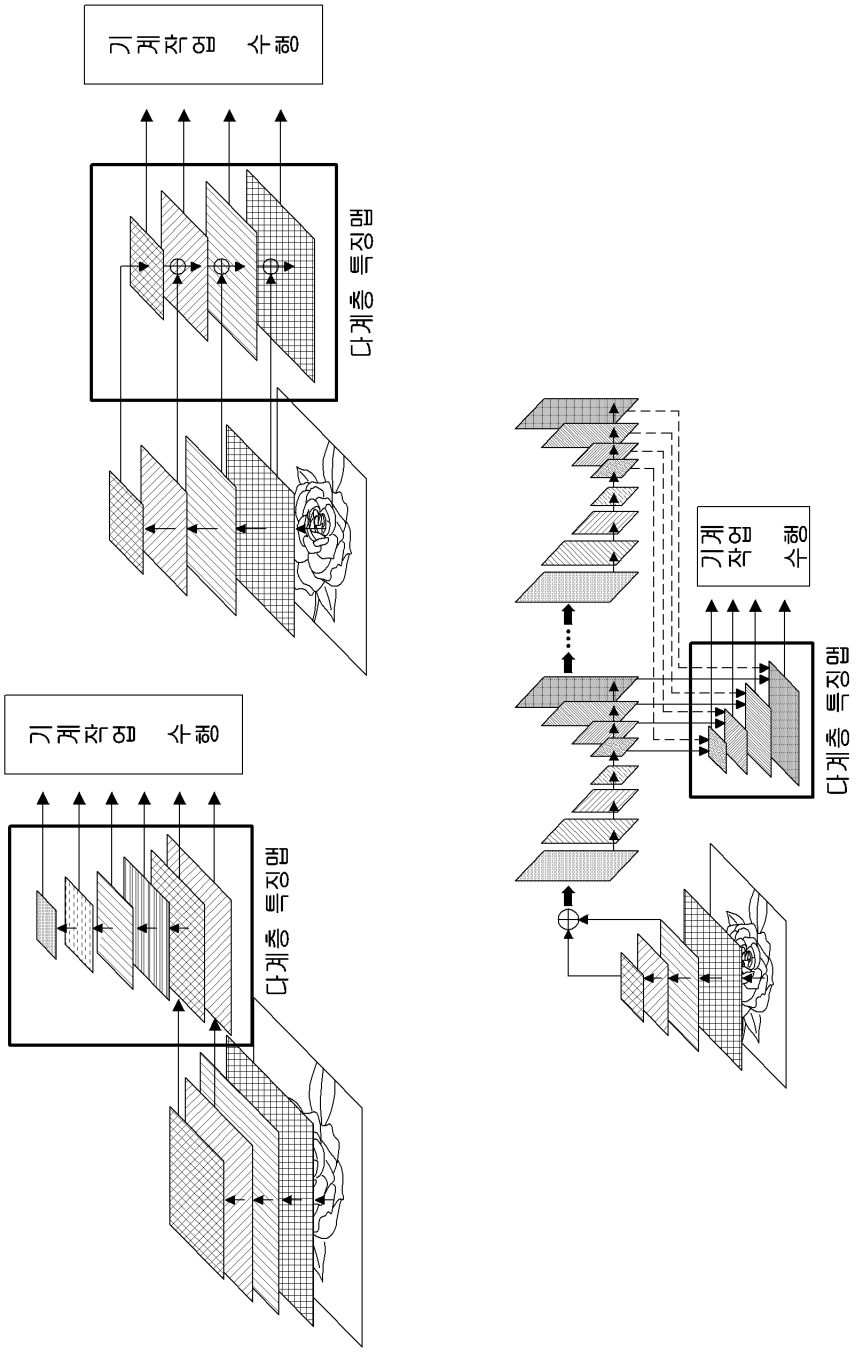
[0364] 따라서, 본 개시는 이하의 특허청구범위 내에 속하는 모든 다른 교체, 수정 및 변경을 포함한다고 할 것이다.

도면

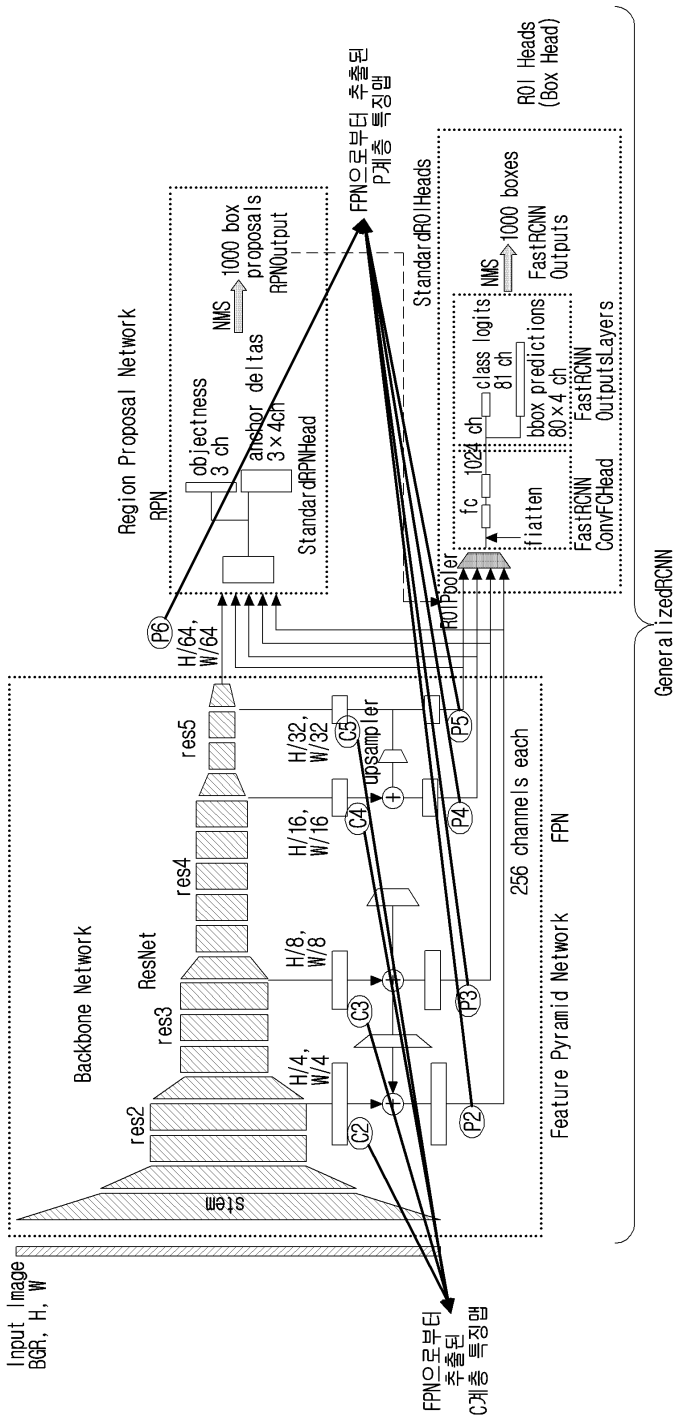
도면1



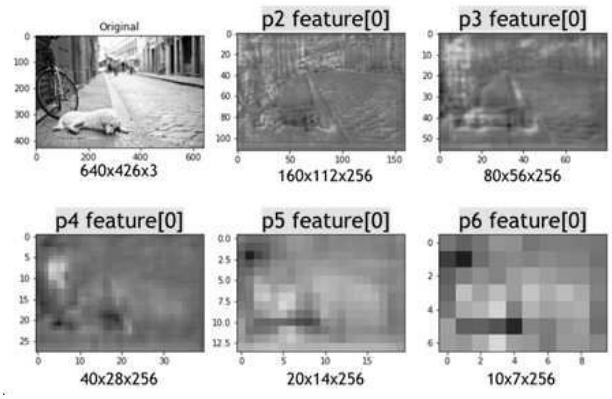
도면2



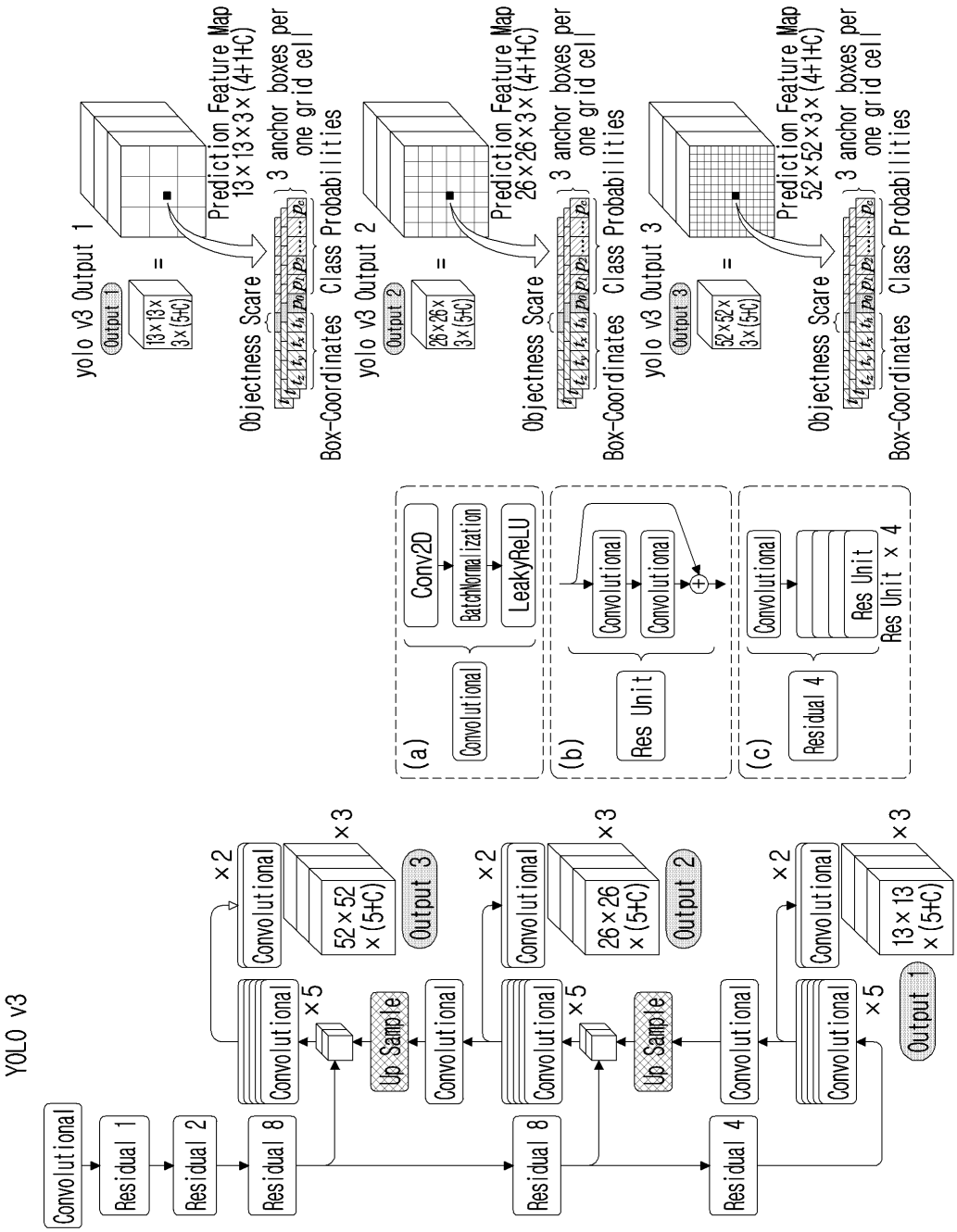
도면3


Input Image
BGR, H, W
Backbone Network
ResNet
stem
res2
res3
res4
res5
H/4, W/4
H/8, W/8
H/16, W/16
H/32, W/32
C2
C3
C4
C5
upsampler
Feature Pyramid Network
FPN
256 channels each
P2
P3
P4
P5
P6
H/64, W/64
C계층 특징맵
추출된
FPN으로부터
Region Proposal Network
RPN
objectness 3 ch
anchor deltas 3×4ch
StandardRPNHead
NMS
1000 box proposals
RPNOutput
P계층 특징맵
FPN으로부터 추출된
StandardROIHeads
ROIPooler
fc 1024 ch
flatten
FastRCNN ConvFCHead
class logits: 81 ch
bbox predictions 80×4 ch
FastRCNN OutputsLayers
NMS
1000 boxes
FastRCNN Outputs
ROI Heads (Box Head)
GeneralizedRCNN

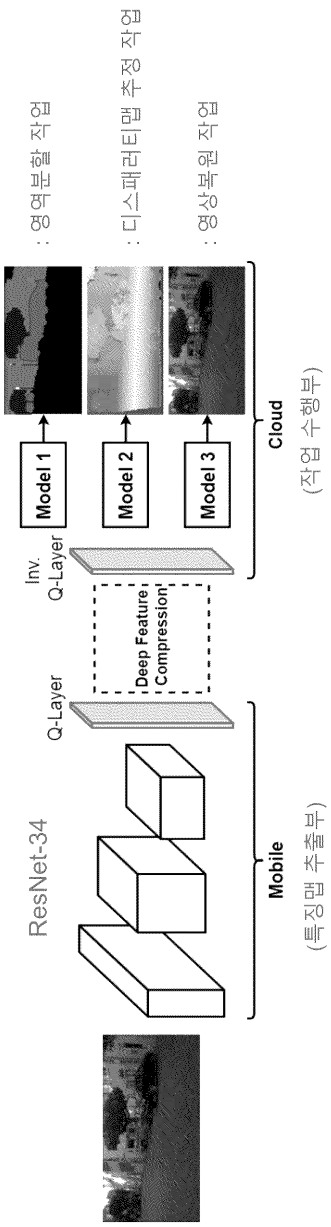
도면4


Original
640x426x3
p2 feature[0]
160x112x256
p3 feature[0]
80x56x256
p4 feature[0]
40x28x256
p5 feature[0]
20x14x256
p6 feature[0]
10x7x256

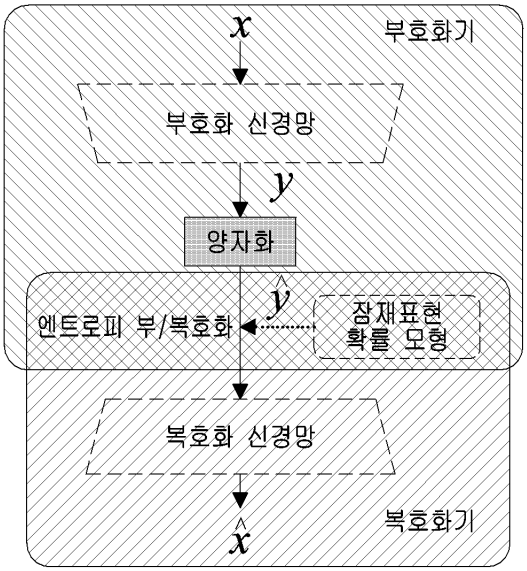
도면5


YOLO v3
Convolutional
Residual 1
Residual 2
Residual 8
Residual 8
Residual 4
Convolutional
Convolutional
×5
Up Sample
Convolutional
Convolutional
Convolutional
×2
Convolutional
52×52
×(5+C)
×3
Output 3
Convolutional
Convolutional
×5
Up Sample
Convolutional
Convolutional
Convolutional
×2
Convolutional
26×26
×(5+C)
×3
Output 2
Convolutional
Convolutional
×5
Convolutional
Convolutional
Convolutional
×2
Convolutional
13×13
×(5+C)
×3
Output 1
(a)
Convolutional
Conv2D
BatchNormalization
LeakyReLU
(b)
Res Unit
Convolutional
Convolutional
(c)
Residual 4
Convolutional
Res Unit
Res Unit × 4
yolo v3 Output 1
Output 1
=
13×13×
3×(5+C)
Prediction Feature Map
13×13×3×(4+1+C)
Objectness Scare
3 anchor boxes per
one grid cell
Box-Coordinates
Class Probabilities
yolo v3 Output 2
Output 2
=
26×26×
3×(5+C)
Prediction Feature Map
26×26×3×(4+1+C)
Objectness Scare
3 anchor boxes per
one grid cell
Box-Coordinates
Class Probabilities
yolo v3 Output 3
Output 3
=
52×52×
3×(5+C)
Prediction Feature Map
52×52×3×(4+1+C)
Objectness Scare
3 anchor boxes per
one grid cell
Box-Coordinates
Class Probabilities

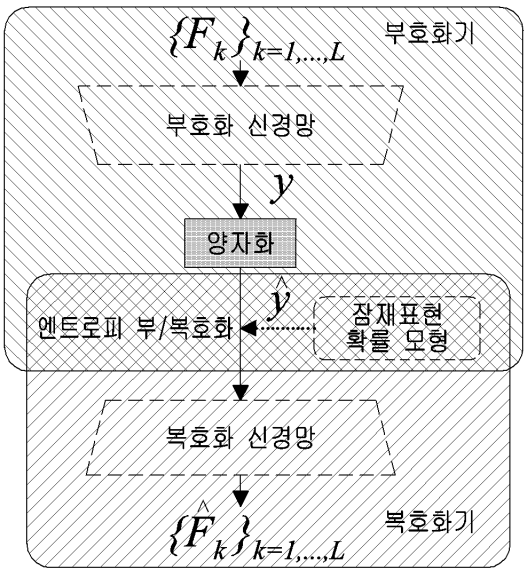
도면6


: 영역분할 작업
: 디스패러티맵 추정 작업
: 영상복원 작업
Model 1
Model 2
Model 3
Cloud
(작업 수행부)
Inv.
Q-Layer
Deep Feature
Compression
Q-Layer
ResNet-34
Mobile
(특징맵 추출부)

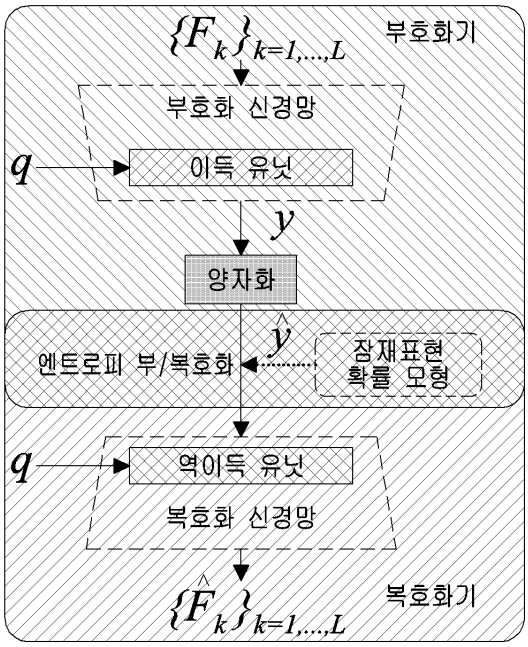
도면7


x
부호화기
부호화 신경망
y
양자화
$\hat{y}$
엔트로피 부/복호화
잠재표현
확률 모형
복호화 신경망
$\hat{x}$
복호화기

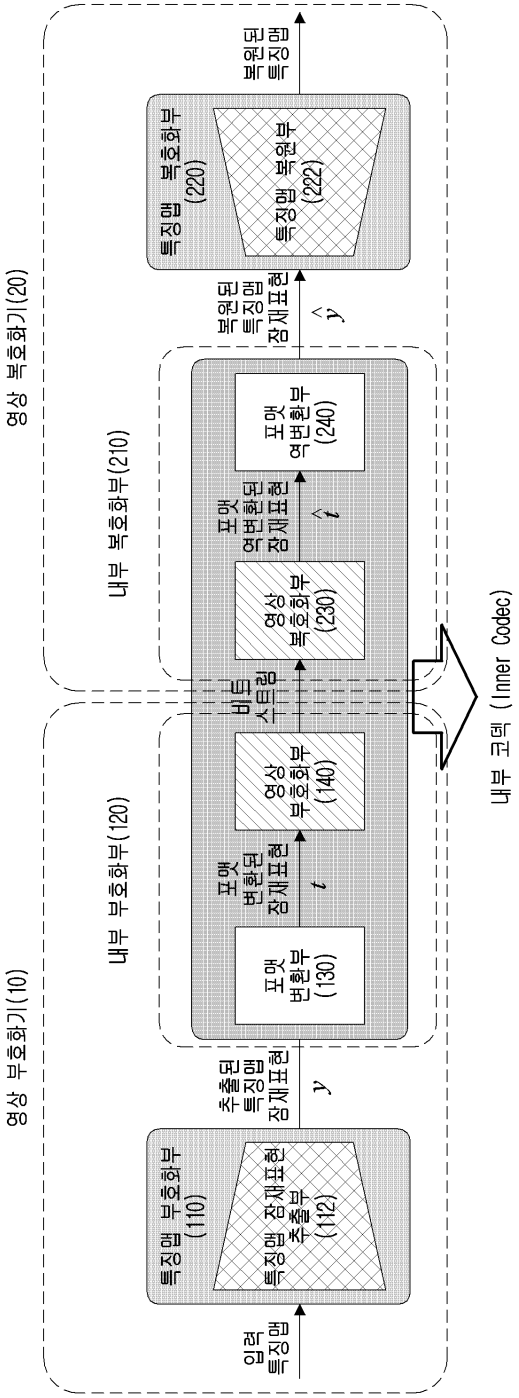
도면8


$\{F_k\}_{k=1,\ldots,L}$
부호화기
부호화 신경망
y
양자화
$\hat{y}$
엔트로피 부/복호화
잠재표현
확률 모형
복호화 신경망
$\{\hat{F}_k\}_{k=1,\ldots,L}$
복호화기

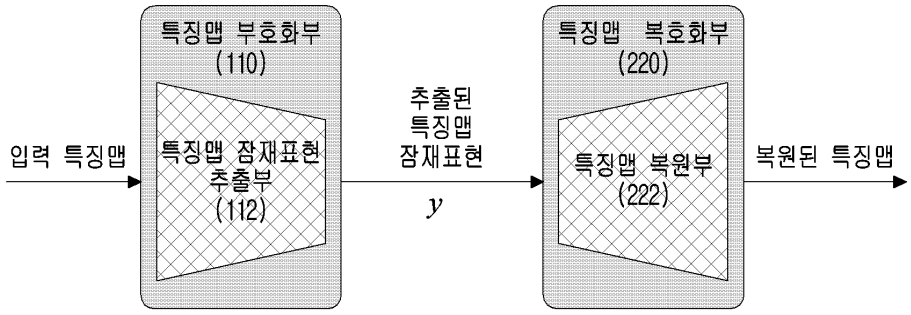
도면9


$\{F_k\}_{k=1,\ldots,L}$
부호화기
부호화 신경망
q
이득 유닛
y
양자화
$\hat{y}$
엔트로피 부/복호화
잠재표현
확률 모형
q
역이득 유닛
복호화 신경망
$\{\hat{F}_k\}_{k=1,\ldots,L}$
복호화기

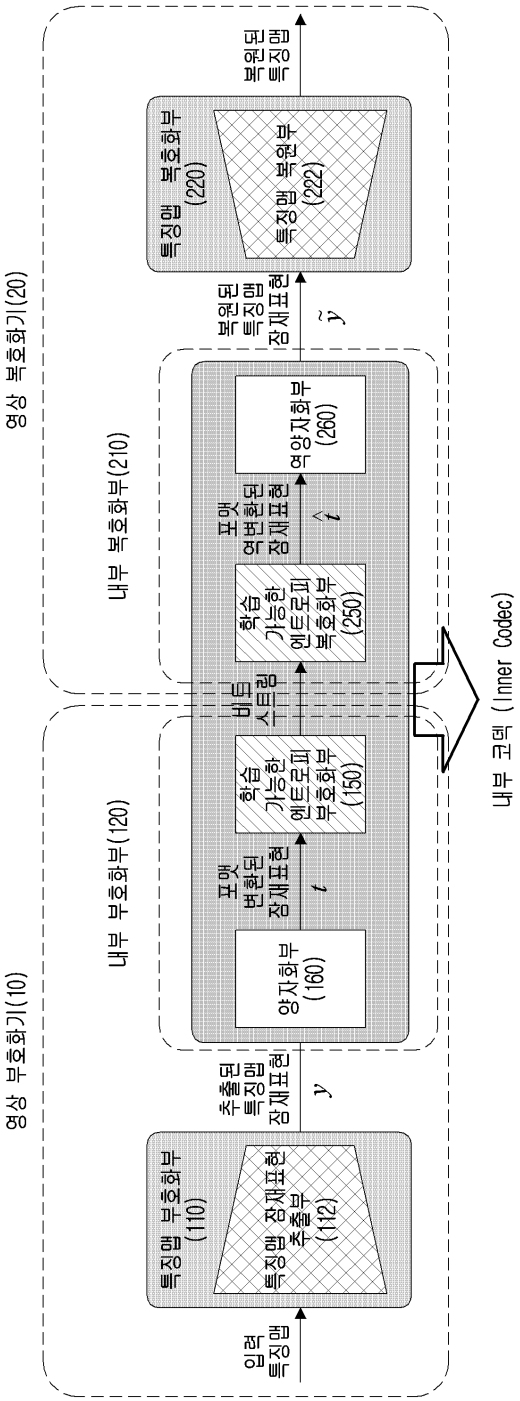
도면10


영상 부호화기(10)
내부 부호화부(120)
특징맵 부호화부 (110)
특징맵 잠재표현 추출부 (112)
입력 특징맵
추출된 특징맵 잠재표현
포맷 변환부 (130)
포맷 변환된 잠재표현
영상 부호화부 (140)
비트 스트림
영상 복호화기(20)
내부 복호화부(210)
영상 복호화부 (230)
포맷 역변환된 잠재표현
포맷 역변환부 (240)
복원된 특징맵 잠재표현
특징맵 복호화부 (220)
특징맵 복원부 (222)
복원된 특징맵
내부 코덱 (Inner Codec)

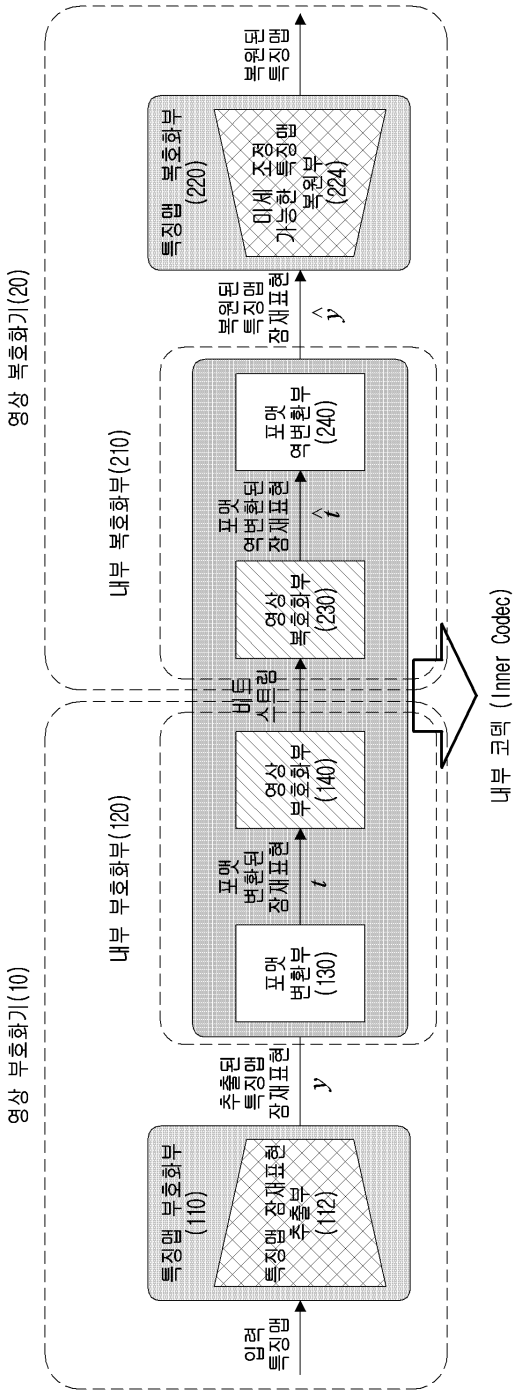
도면11


특징맵 부호화부
(110)
입력 특징맵
특징맵 잠재표현
추출부
(112)
추출된
특징맵
잠재표현
y
특징맵 복호화부
(220)
특징맵 복원부
(222)
복원된 특징맵

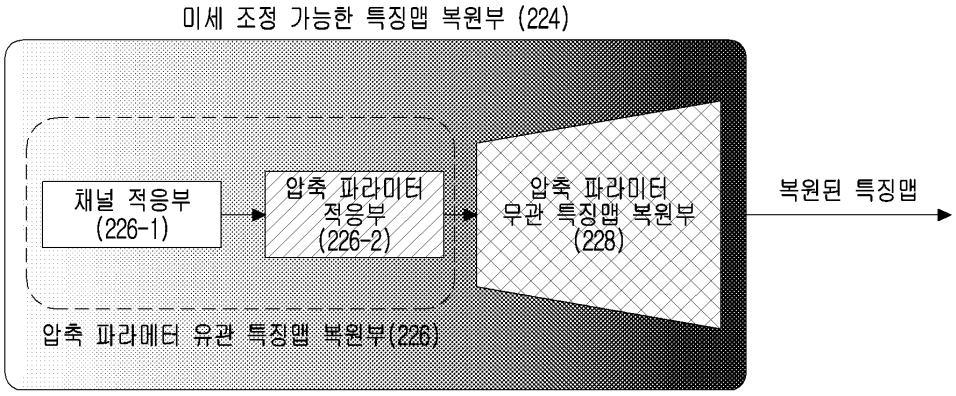
도면12


영상 부호화기(10)
영상 복호화기(20)
내부 부호화부(120)
내부 복호화부(210)
입력 특징맵
특징맵 부호화부 (110)
특징맵 잠재표현 추출부 (112)
추출된 특징맵 잠재표현
양자화부 (160)
포맷 변환된 잠재표현
학습 가능한 엔트로피 부호화부 (150)
비트 스트림
학습 가능한 엔트로피 복호화부 (250)
포맷 역변환된 잠재표현
역양자화부 (260)
복원된 특징맵 잠재표현
특징맵 복호화부 (220)
특징맵 복원부 (222)
복원된 특징맵
내부 코덱 (Inner Codec)

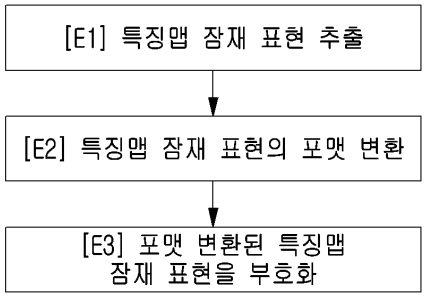
도면13


영상 부호화기(10)
내부 부호화부(120)
특징맵 부호화부 (110)
특징맵 잠재표현 추출부 (112)
추출된 특징맵 잠재표현
y
입력 특징맵
포맷 변환부 (130)
포맷 변환된 잠재표현
t
영상 부호화부 (140)
비트 스트림
내부 코덱 (Inner Codec)
영상 복호화기(20)
내부 복호화부(210)
영상 복호화부 (230)
포맷 역변환된 잠재표현
포맷 역변환부 (240)
복원된 특징맵 잠재표현
특징맵 복호화부 (220)
미세 조정형 가능한 특징맵 복원부 (224)
복원된 특징맵

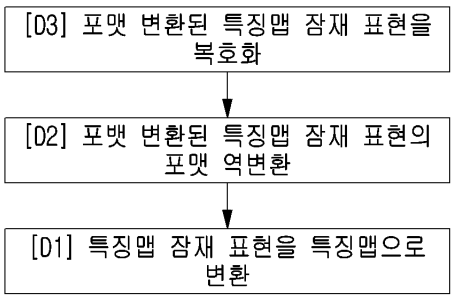
도면14



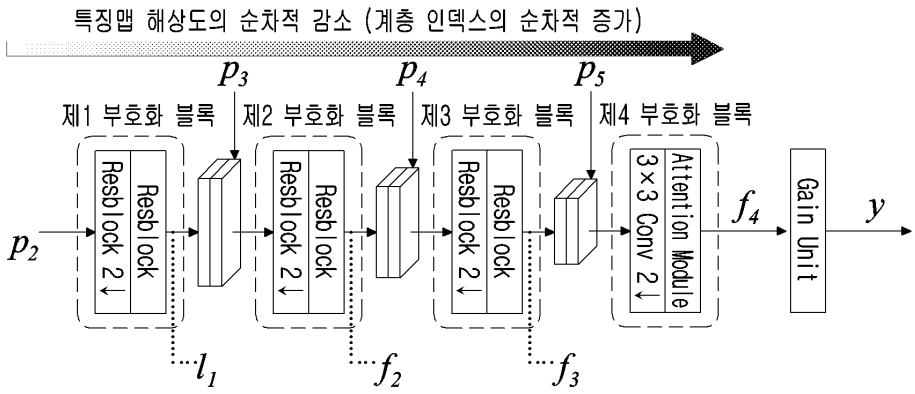
도면15



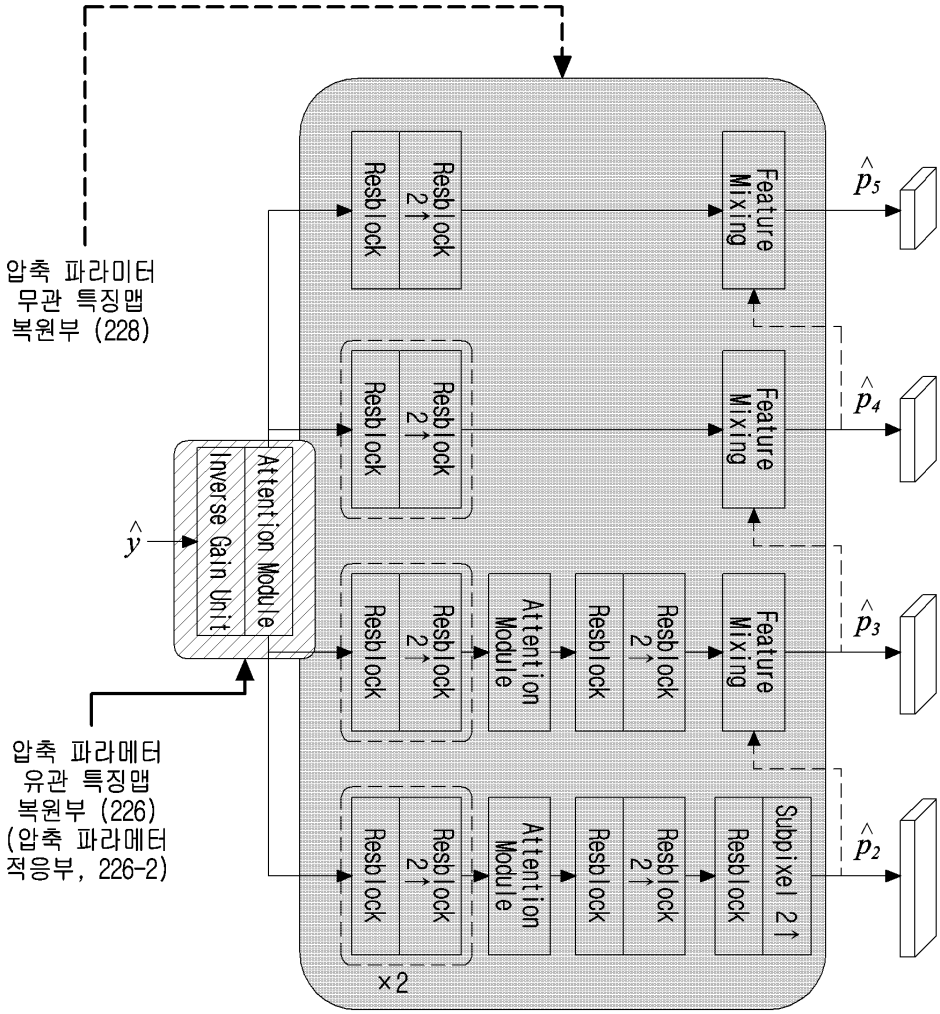
도면16



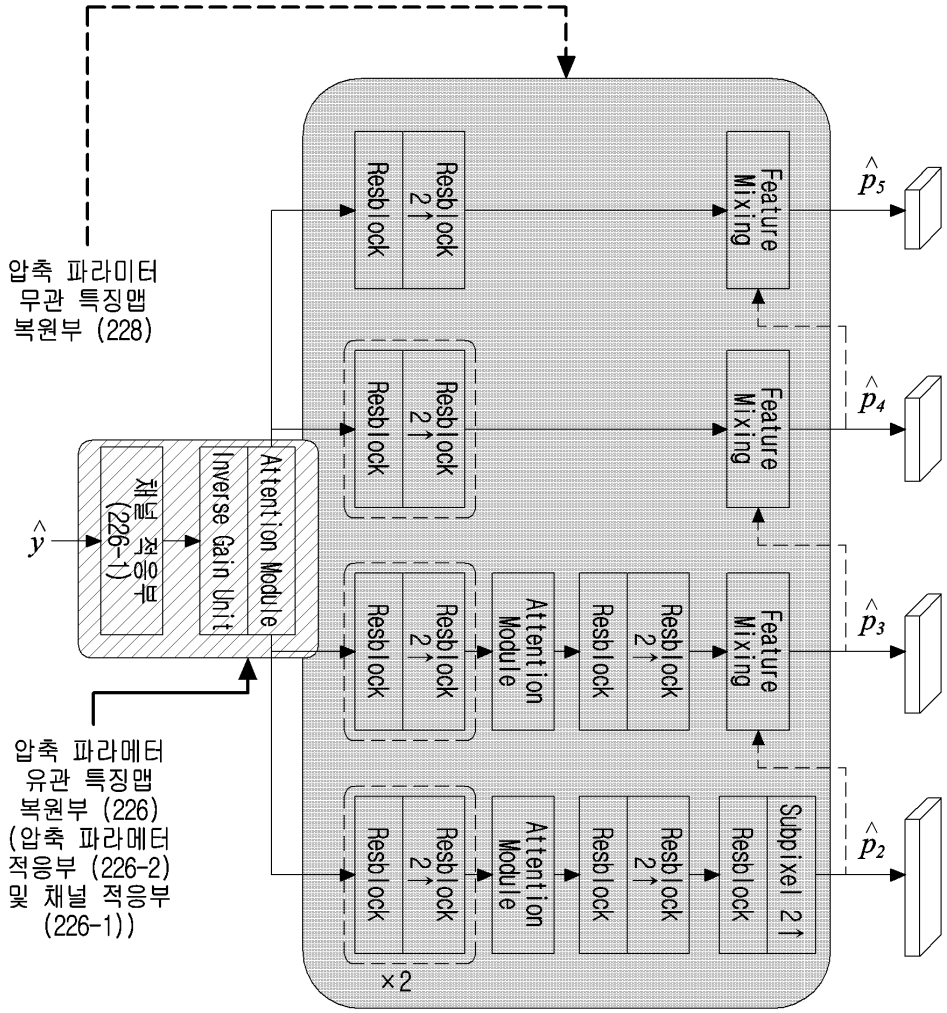
도면17


특징맵 해상도의 순차적 감소 (계층 인덱스의 순차적 증가)
p_3
p_4
p_5
제1 부호화 블록
제2 부호화 블록
제3 부호화 블록
제4 부호화 블록
Resblock 2↓
Resblock
Attention Module
3×3 Conv 2↓
Gain Unit
p_2
f_4
y
l_1
f_2
f_3

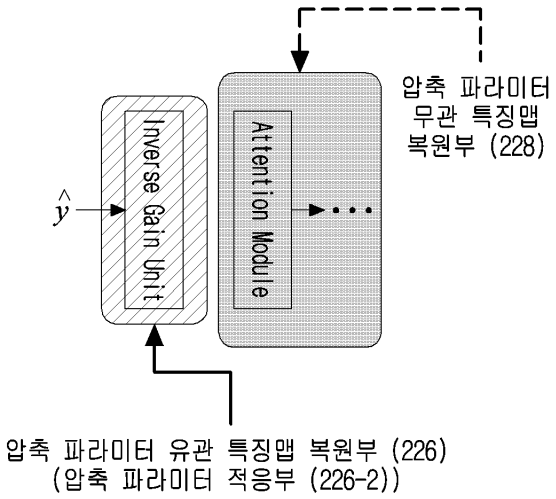
도면18


압축 파라미터 무관 특징맵 복원부 (228)
압축 파라메터 유관 특징맵 복원부 (226) (압축 파라메터 적응부, 226-2)
Inverse Gain Unit
Attention Module
$\hat{y}$
Resblock
Resblock 2↑
Attention Module
Feature Mixing
Subpixel 2↑
$\hat{p}_5$
$\hat{p}_4$
$\hat{p}_3$
$\hat{p}_2$
×2

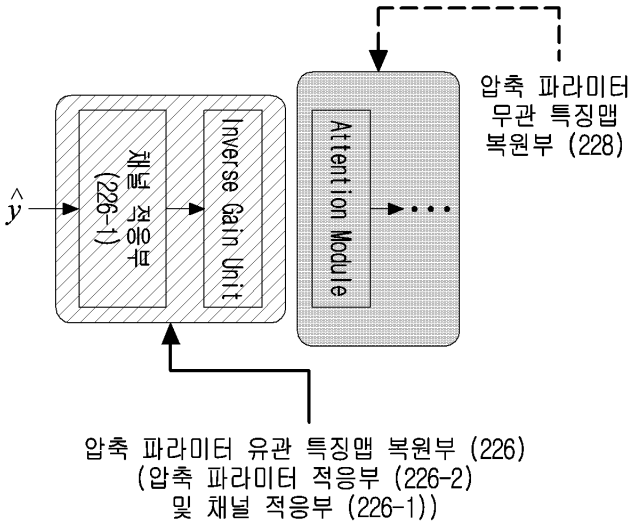
도면19


압축 파라미터
무관 특징맵
복원부 (228)
Resblock
Resblock 2↑
Feature Mixing
Attention Module
Inverse Gain Unit
채널 적응부 (226-1)
Subpixel 2↑
×2
압축 파라메터
유관 특징맵
복원부 (226)
(압축 파라메터
적응부 (226-2)
및 채널 적응부
(226-1))

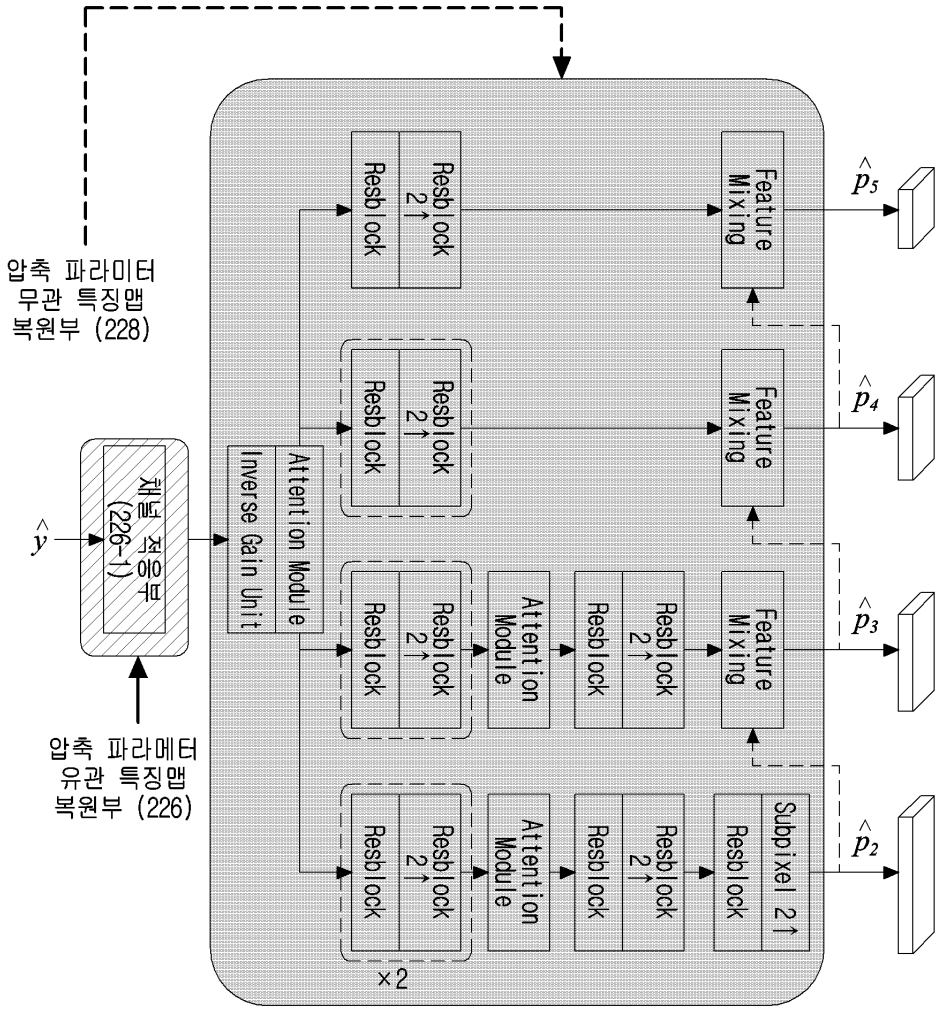
도면20


Inverse Gain Unit
Attention Module
압축 파라미터
무관 특징맵
복원부 (228)
압축 파라미터 유관 특징맵 복원부 (226)
(압축 파라미터 적응부 (226-2))

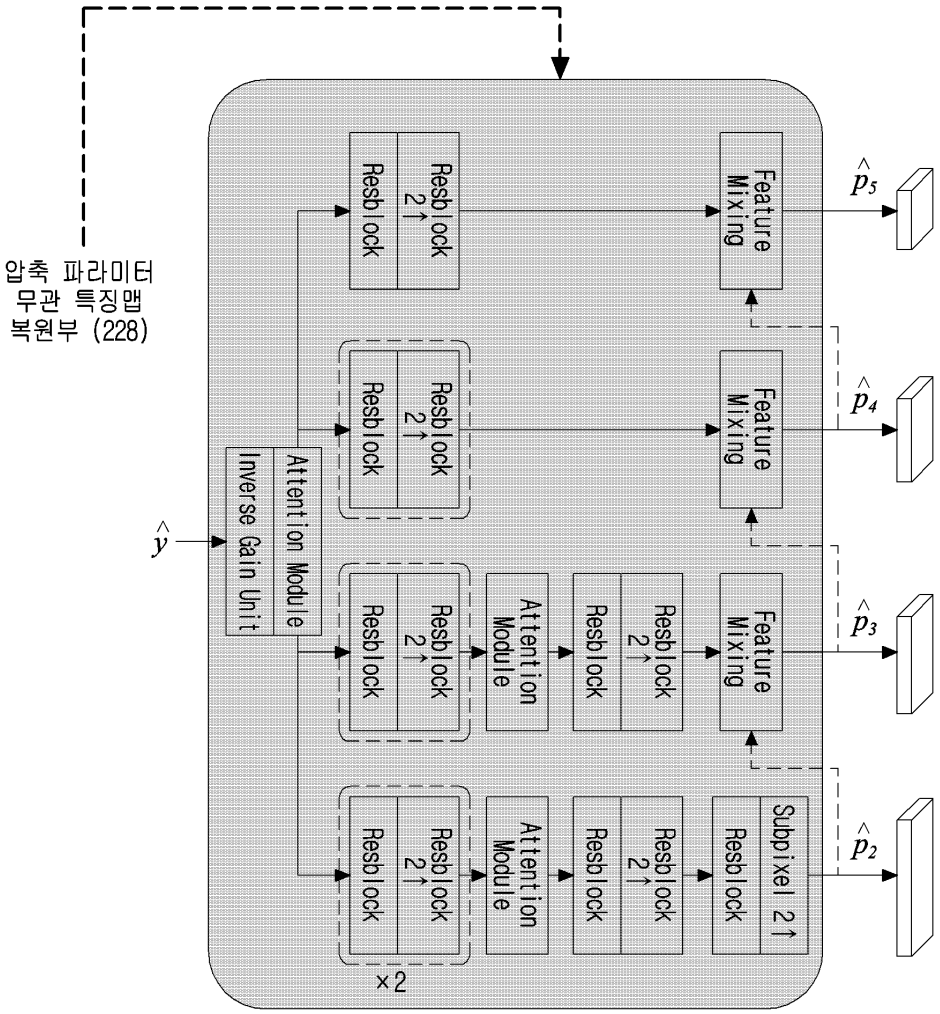
도면21


압축 파라미터
무관 특징맵
복원부 (228)
Attention Module
Inverse Gain Unit
채널 적응부
(226-1)
압축 파라미터 유관 특징맵 복원부 (226)
(압축 파라미터 적응부 (226-2)
및 채널 적응부 (226-1))

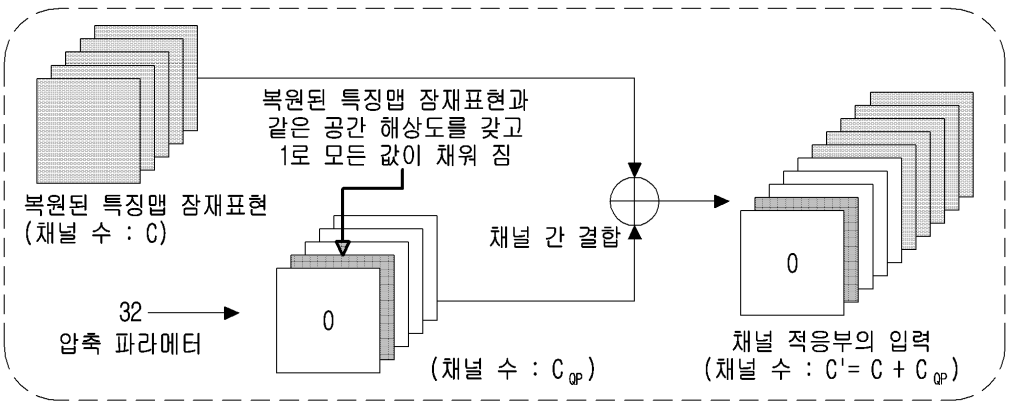
도면22


압축 파라미터
무관 특징맵
복원부 (228)
채널 적응부
(226-1)
Attention Module
Inverse Gain Unit
압축 파라메터
유관 특징맵
복원부 (226)
Resblock
Resblock 2↑
Attention Module
Feature Mixing
Subpixel 2↑
×2

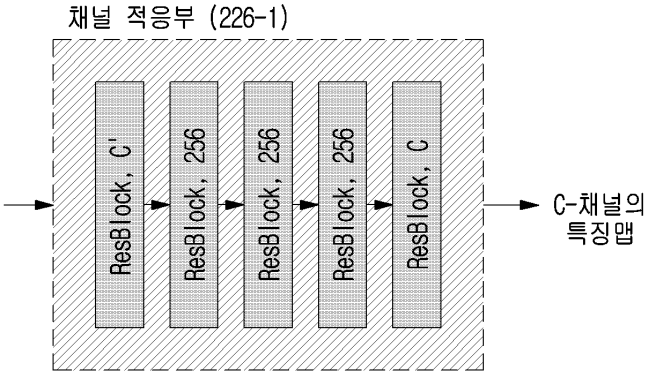
도면23

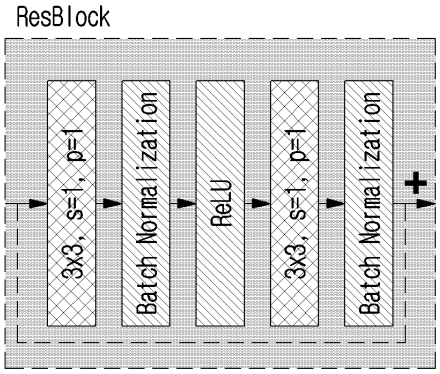

압축 파라미터
무관 특징맵
복원부 (228)
ŷ
Inverse Gain Unit
Attention Module
Resblock
Resblock 2↑
Attention Module
Feature Mixing
Subpixel 2↑
×2
p̂5
p̂4
p̂3
p̂2

도면24

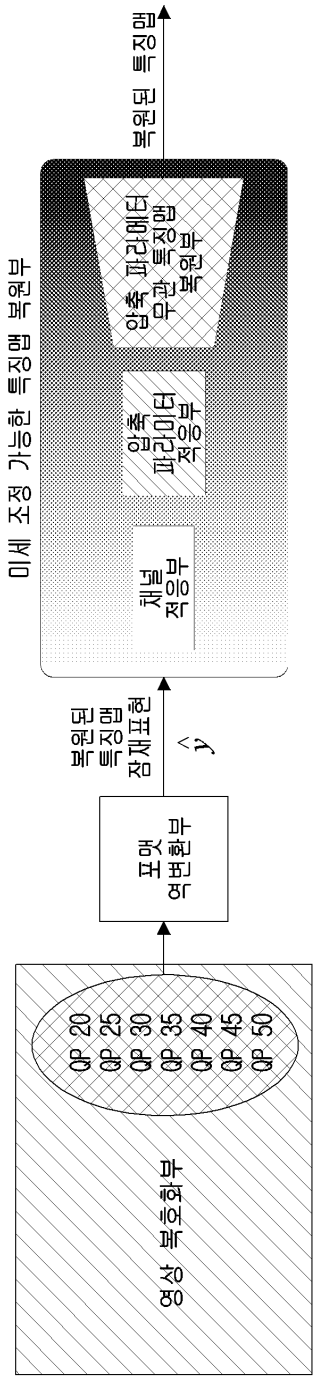

복원된 특징맵 잠재표현
(채널 수 : C)
복원된 특징맵 잠재표현과
같은 공간 해상도를 갖고
1로 모든 값이 채워 짐
32
압축 파라메터
0
(채널 수 : C_QP)
채널 간 결합
0
채널 적응부의 입력
(채널 수 : C'= C + C_QP)

도면25

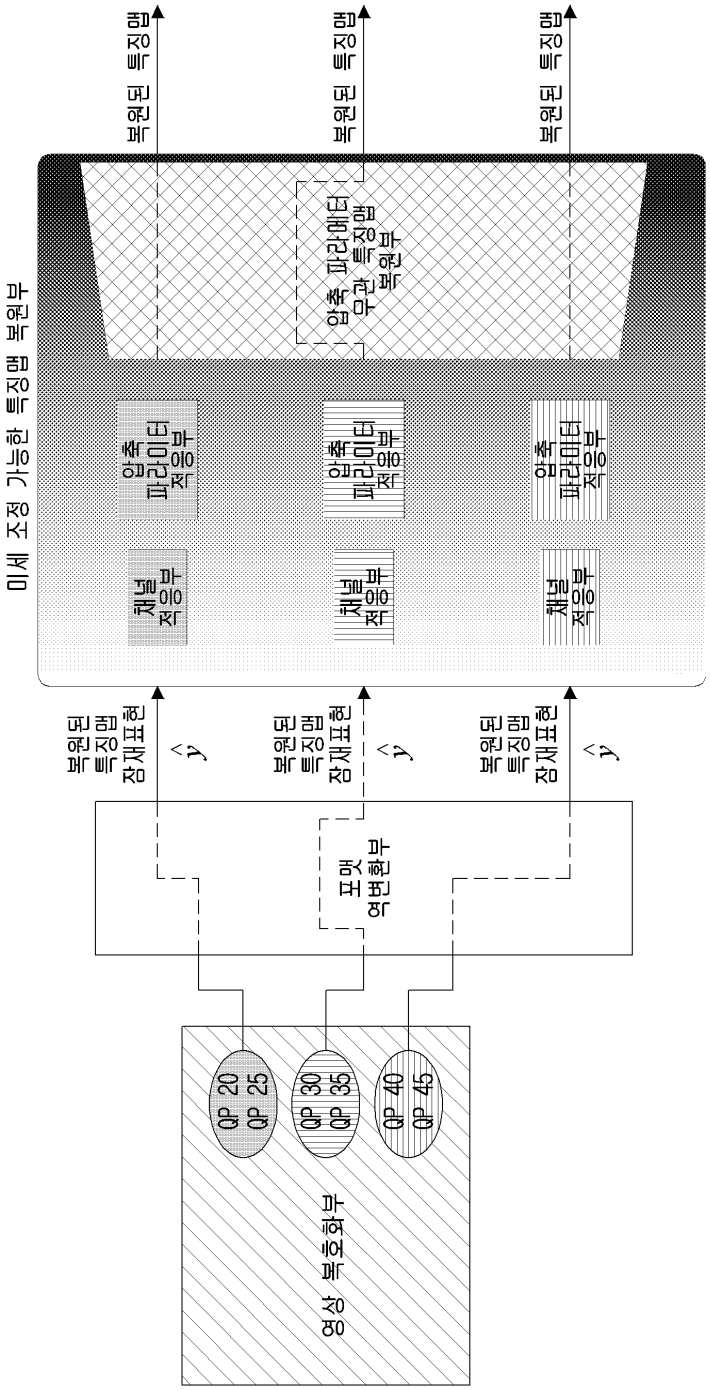

채널 적응부 (226-1)
ResBlock, C'
ResBlock, 256
ResBlock, 256
ResBlock, 256
ResBlock, C
C-채널의
특징맵


ResBlock
3x3, s=1, p=1
Batch Normalization
ReLU
3x3, s=1, p=1
Batch Normalization

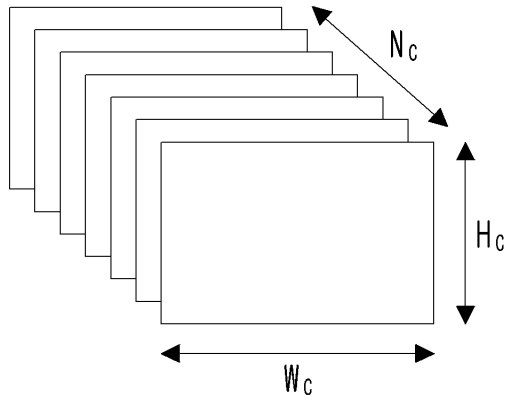
도면26


복원된 특징맵
미세 조정 가능한 특징맵 복원부
압축 파라메터 무관 특징맵 복원부
압축 파라미터 적응부
채널 적응부
복원된 특징맵 잠재표현
$\hat{y}$
엔트로피 복호화부
영상 복호화부
QP 20
QP 25
QP 30
QP 35
QP 40
QP 45
QP 50

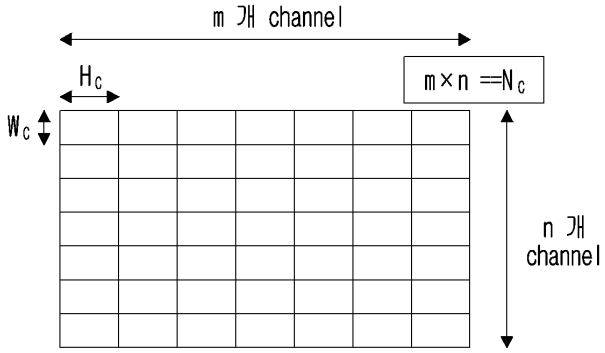
도면27


복원된 특징맵
복원된 특징맵
복원된 특징맵
미세 조정 가능한 특징맵 복원부
압축 파라메터 무관 특징맵 복원부
압축 파라미터 적응부
압축 파라미터 적응부
압축 파라미터 적응부
채널 적응부
채널 적응부
채널 적응부
복원된 특징맵 잠재표현
복원된 특징맵 잠재표현
복원된 특징맵 잠재표현
포맷 역변환부
QP 20
QP 25
QP 30
QP 35
QP 40
QP 45
영상 복호화부

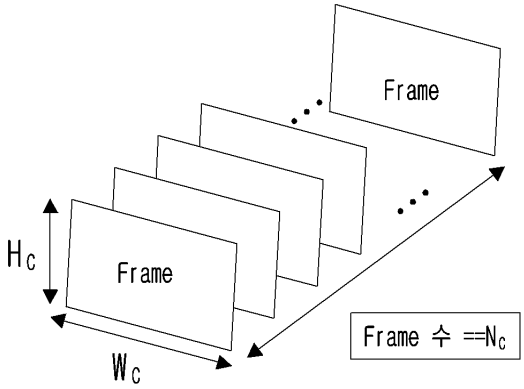
도면28



도면29



도면30



도면31

