(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0001454
(43) 공개일자 2025년01월06일

(51) 국제특허분류(Int. Cl.)
H04N 19/30 (2014.01) *G06N 3/0464* (2023.01)
G06N 3/048 (2023.01) *H04N 19/136* (2014.01)
H04N 19/176 (2014.01)
(52) CPC특허분류
H04N 19/30 (2015.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2024-0084730
(22) 출원일자 2024년06월27일
심사청구일자 없음
(30) 우선권주장
1020230083810 2023년06월28일 대한민국(KR)
(뒷면에 계속)

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

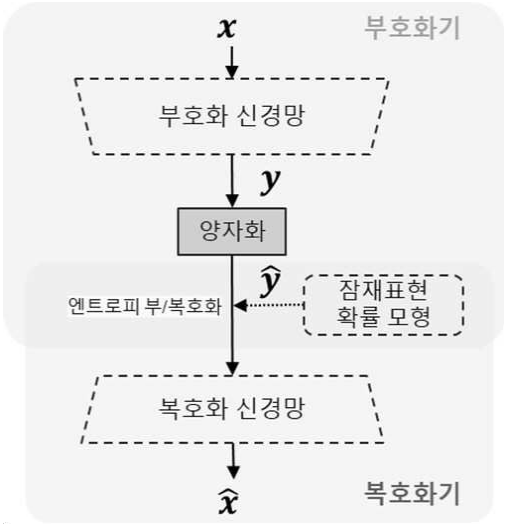
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **신경망 기반의 다계층 특징맵 압축 방법 및 장치**

(57) 요 약

본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체는, 입력 영상으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵을 추출하는 단계, 상기 다계층 특징맵을 연속된 부호화 블록들을 통해 순차적으로 부호화하여 단일 계층의 융합 잠재표현을 출력하는 단계, 및 상기 단일 계층의 융합 잠재표현을 비트스트림으로 부호화하는 단계를 포함할 수 있다.

대 표 도 - 도7



(52) CPC특허분류
G06N 3/048 (2023.01)
H04N 19/136 (2015.01)
H04N 19/176 (2015.01)

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

(30) 우선권주장
1020230089970 2023년07월11일 대한민국(KR)
1020230134679 2023년10월10일 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

입력 영상으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵을 추출하는 단계;

상기 다계층 특징맵을 연속된 부호화 블록들을 통해 순차적으로 부호화하여 단일 계층의 융합 잠재표현을 출력하는 단계; 및

상기 단일 계층의 융합 잠재표현을 비트스트림으로 부호화하는 단계를 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 2

제1항에 있어서,

상기 연속된 부호화 블록들은, 상기 다계층 특징맵의 가장 높은 해상도를 가지는 특징맵을 입력으로 하는 제1 부호화 블록, 제1 부호화 블록의 출력인 제1 중간 잠재표현을 입력으로 하는 제2 부호화 블록, 제2 부호화 블록의 출력인 제1 융합 잠재표현을 입력으로 하는 제3 부호화 블록, 및 제3 부호화 블록의 출력인 제2 융합 잠재표현을 입력으로 하여 상기 단일 계층의 융합 잠재표현을 출력하는 제4 부호화 블록을 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 3

제2항에 있어서,

상기 제2 부호화 블록, 상기 제3 부호화 블록, 및 상기 제4 부호화 블록의 입력은, 입력된 중간 잠재표현 또는 융합 잠재표현과 동일한 해상도를 가지는 상기 다계층 특징맵의 특징맵을 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 4

제2항에 있어서,

상기 제1 부호화 블록, 상기 제2 부호화 블록, 및 상기 제3 부호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층을 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 5

제4항에 있어서,

상기 제4 부호화 블록은, 상기 LeakyReLU의 비선형층을 포함하지 않는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 6

제2항에 있어서,

상기 제1 부호화 블록, 상기 제2 부호화 블록, 상기 제3 부호화 블록 및 상기 제4 부호화 블록의 출력의 해상도는 입력의 해상도보다 낮은 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 7

제1항에 있어서,

상기 연속된 부호화 블록들은, 상기 다계층 특징맵의 가장 높은 해상도를 가지는 특징맵을 입력으로 하는 제1 부호화 블록, 제1 부호화 블록의 출력인 제1 중간 잠재표현을 입력으로 하는 제2 부호화 블록, 제2 부호화 블록

의 출력인 제1 융합 잠재표현보다 해상도가 낮은 상기 다계층 특징맵의 특징맵을 입력으로 하는 제3 부호화 블록, 및 제3 부호화 블록의 출력인 제2 중간 잠재표현 및 상기 제1 융합 잠재표현을 입력으로 하여 상기 단일 계층의 융합 잠재표현을 출력하는 제4 부호화 블록을 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 8

제7항에 있어서,

상기 제2 부호화 블록 및 상기 제4 부호화 블록의 입력은, 입력된 중간 잠재표현 또는 융합 잠재표현과 동일한 해상도를 가지는 상기 다계층 특징맵의 특징맵을 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 9

제7항에 있어서,

상기 제1 부호화 블록, 상기 제2 부호화 블록, 및 상기 제3 부호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층을 포함하는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 10

제9항에 있어서,

상기 제4 부호화 블록은, 상기 LeakyReLU의 비선형층을 포함하지 않는 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 11

제7항에 있어서,

상기 제1 부호화 블록, 상기 제2 부호화 블록, 및 상기 제4 부호화 블록의 출력의 해상도는 입력의 해상도보다 낮고,

상기 제3 부호화 블록의 출력의 해상도는 입력의 해상도보다 높은 것을 특징으로 하는, 다계층 특징맵 부호화 방법.

청구항 12

비트스트림으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵에 대한 단일 계층의 융합 잠재표현을 복호화하는 단계; 및

상기 단일 계층의 융합 잠재표현으로부터 연속된 복호화 블록들을 통해 상기 다계층 특징맵을 순차적으로 복호화하는 단계를 포함하는 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 13

제12항에 있어서,

상기 연속된 복호화 블록들은, 상기 단일 계층의 융합 잠재표현을 입력으로 하는 제1 복호화 블록, 제1 복호화 블록의 출력인 가장 낮은 해상도를 가지는 상기 다계층 특징맵의 제1 특징맵을 입력으로 하는 제2 복호화 블록, 상기 제1 특징맵보다 높은 해상도를 가지는 제2 복호화 블록의 출력인 제2 특징맵을 입력으로 하는 제3 복호화 블록, 및 상기 제2 특징맵보다 높은 해상도를 가지는 제3 복호화 블록의 출력인 제3 특징맵을 입력으로 하여 가장 큰 해상도를 가지는 상기 다계층 특징맵의 제4 특징맵을 출력하는 제4 복호화 블록을 포함하는 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 14

제13항에 있어서,

상기 제1 복호화 블록, 상기 제2 복호화 블록, 및 상기 제3 복호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층 포함하는 것을 특징으로 하는, 다계층 특징맵 복호화

방법.

청구항 15

제14항에 있어서,

상기 제4 복호화 블록은, 상기 LeakyReLU의 비선형층을 포함하지 않는 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 16

제13항에 있어서,

상기 제1 복호화 블록, 상기 제2 복호화 블록, 상기 제3 복호화 블록 및 상기 제4 복호화 블록의 출력의 해상도는 입력의 해상도보다 높은 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 17

제12항에 있어서,

상기 연속된 복호화 블록들은, 상기 단일 계층의 융합 잠재표현을 입력으로 하는 제1 복호화 블록, 제1 복호화 블록의 출력인 제1 특징맵을 입력으로 하여 가장 낮은 해상도를 가지는 상기 다계층 특징맵의 제2 특징맵을 출력하는 제2 복호화 블록, 상기 제1 특징맵을 입력으로 하여 상기 제1 특징맵보다 높은 해상도를 가지는 제3 특징맵을 출력하는 제3 복호화 블록, 상기 제3 특징맵을 입력으로 가장 큰 해상도를 가지는 상기 다계층 특징맵의 제4 특징맵을 출력하는 제4 복호화 블록을 포함하는 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 18

제17항에 있어서,

상기 제2 복호화 블록, 상기 제3 복호화 블록, 및 상기 제4 복호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층 포함하는 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 19

제17항에 있어서,

상기 제1 복호화 블록, 상기 제3 복호화 블록, 및 상기 제4 복호화 블록의 출력의 해상도는 입력의 해상도보다 높고,

상기 제2 복호화 블록의 출력의 해상도는 입력의 해상도보다 낮은 것을 특징으로 하는, 다계층 특징맵 복호화 방법.

청구항 20

다계층 특징맵 부호화 방법에 의해 생성된 비트스트림을 저장하는 컴퓨터 판독가능한 기록매체에 있어서,

상기 다계층 특징맵 부호화 방법은, 입력 영상으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵을 추출하는 단계;

상기 다계층 특징맵을 연속된 부호화 블록들을 통해 순차적으로 부호화하여 단일 계층의 융합 잠재표현을 출력하는 단계; 및

상기 단일 계층의 융합 잠재표현을 비트스트림으로 부호화하는 단계를 포함하는 것을 특징으로 하는, 컴퓨터 판독가능한 기록매체.

발명의 설명

기 술 분 야

[0001] 본 개시는 인공신경망을 통해 추출된 특징맵의 부호화 방법을 제공함으로써 특징맵의 압축 비트량을 줄이면서도

복호화된 특징맵을 이용하여 수행되는 기계작업의 성능 저하는 최소화하기 위한 기술에 관한 것이다.

배 경 기 술

[0002] 기계작업이 대형 서버 뿐 아니라, 모바일 장치를 포함한 다양한 장치에서 광범위하게 사용되면서 특징맵 추출 수단과 작업 수행 수단이 같은 장치 내에 존재하지 않고 서로 다른 장치 내에 존재하는 경우도 점점 많아지고 있다.

[0003] 특징맵 추출 수단과 작업 수행 수단이 분리되어 있는 경우 추출된 특징맵을 작업 수행 수단에 전달해야만 하나 특징맵의 데이터 양이 매우 크므로 작업 수행 성능의 저하를 최소화하면서도 특징맵의 데이터 양을 획기적으로 줄이기 위한 특징맵 부호화 방법이 필요하다.

발명의 내용

해결하려는 과제

[0004] 종래기술은 다계층 특징맵을 독립적으로 부호화하거나 하나의 융합 특징맵으로 만들어 부호화를 수행하며, 전자의 경우 병렬적으로 여러 번 부호화 함에 따른 복잡도 증가 문제가 있고, 후자의 경우 융합과정과 잠재표현 추출 과정이 직렬적으로 수행됨에 따른 복잡도 증가 문제와 계층간 순차적인 속성을 충분히 활용하지 못한다는 단점이 있다.

과제의 해결 수단

[0005] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체는, 입력 영상으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵을 추출하는 단계, 상기 다계층 특징맵을 연속된 부호화 블록들을 통해 순차적으로 부호화하여 단일 계층의 융합 잠재표현을 출력하는 단계, 및 상기 단일 계층의 융합 잠재표현을 비트스트림으로 부호화하는 단계를 포함할 수 있다.

[0006] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 연속된 부호화 블록들은, 상기 다계층 특징맵의 가장 높은 해상도를 가지는 특징맵을 입력으로 하는 제1 부호화 블록, 제1 부호화 블록의 출력인 제1 중간 잠재표현을 입력으로 하는 제2 부호화 블록, 제2 부호화 블록의 출력인 제1 융합 잠재표현을 입력으로 하는 제3 부호화 블록, 및 제3 부호화 블록의 출력인 제2 융합 잠재표현을 입력으로 하여 상기 단일 계층의 융합 잠재표현을 출력하는 제4 부호화 블록을 포함할 수 있다.

[0007] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제2 부호화 블록, 상기 제3 부호화 블록, 및 상기 제4 부호화 블록의 입력은, 입력된 중간 잠재표현 또는 융합 잠재표현과 동일한 해상도를 가지는 상기 다계층 특징맵의 특징맵을 포함할 수 있다.

[0008] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제1 부호화 블록, 상기 제2 부호화 블록, 및 상기 제3 부호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층을 포함할 수 있다.

[0009] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제4 부호화 블록은, 상기 LeakyReLU의 비선형층을 포함하지 않을 수 있다.

[0010] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제1 부호화 블록, 상기 제2 부호화 블록, 상기 제3 부호화 블록 및 상기 제4 부호화 블록의 출력의 해상도는 입력의 해상도보다 낮을 수 있다.

[0011] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 연속된 부호화 블록들은, 상기 다계층 특징맵의 가장 높은 해상도를 가지는 특징맵을 입력으로 하는 제1 부호화 블록, 제1 부호화 블록의 출력인 제1 중간 잠재표현을 입력으로 하는 제2 부호화 블록, 제2 부호화 블록의 출력인 제1 융합 잠재표현보다 해상도가 낮은 상기 다계층 특징맵의 특징맵을 입력으로 하는 제3 부호화 블록, 및 제3 부호화 블록의 출력인 제2 중간 잠재표현 및 상기 제1 융합 잠재표현을 입력으로 하여 상기 단일 계층의 융합 잠재표현을 출력하는 제4 부호화 블록을 포함할 수 있다.

[0012] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제2 부호화 블록 및 상기 제4 부호화 블록의 입력은, 입력된 중간 잠재표현 또는 융합 잠재표현과 동일한 해상도를 가지는 상기 다계층 특징맵의 특징맵을 포함할 수 있다.

[0013] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제1 부호화 블록, 상기 제2 부호화 블록, 및 상기 제3 부호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층을 포함할 수 있다.

[0014] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제4 부호화 블록은, 상기 LeakyReLU의 비선형층을 포함하지 않을 수 있다.

[0015] 본 개시의 다계층 특징맵 부호화 방법, 장치 및 기록매체에 있어서, 상기 제1 부호화 블록, 상기 제2 부호화 블록, 및 상기 제4 부호화 블록의 출력의 해상도는 입력의 해상도보다 낮고, 상기 제3 부호화 블록의 출력의 해상도는 입력의 해상도보다 높을 수 있다.

[0016] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체는, 비트스트림으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵에 대한 단일 계층의 융합 잠재표현을 복호화하는 단계; 및 상기 단일 계층의 융합 잠재표현으로부터 연속된 복호화 블록들을 통해 상기 다계층 특징맵을 순차적으로 복호화하는 단계를 포함할 수 있다.

[0017] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 연속된 복호화 블록들은, 상기 단일 계층의 융합 잠재표현을 입력으로 하는 제1 복호화 블록, 제1 복호화 블록의 출력인 가장 낮은 해상도를 가지는 상기 다계층 특징맵의 제1 특징맵을 입력으로 하는 제2 복호화 블록, 상기 제1 특징맵보다 높은 해상도를 가지는 제2 복호화 블록의 출력인 제2 특징맵을 입력으로 하는 제3 복호화 블록, 및 상기 제2 특징맵보다 높은 해상도를 가지는 제3 복호화 블록의 출력인 제3 특징맵을 입력으로 하여 가장 큰 해상도를 가지는 상기 다계층 특징맵의 제4 특징맵을 출력하는 제4 복호화 블록을 포함할 수 있다.

[0018] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 제1 복호화 블록, 상기 제2 복호화 블록, 및 상기 제3 복호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층 포함할 수 있다.

[0019] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 제4 복호화 블록은, 상기 LeakyReLU의 비선형층을 포함하지 않을 수 있다.

[0020] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 제1 복호화 블록, 상기 제2 복호화 블록, 상기 제3 복호화 블록 및 상기 제4 복호화 블록의 출력의 해상도는 입력의 해상도보다 높을 수 있다.

[0021] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 연속된 복호화 블록들은, 상기 단일 계층의 융합 잠재표현을 입력으로 하는 제1 복호화 블록, 제1 복호화 블록의 출력인 제1 특징맵을 입력으로 하여 가장 낮은 해상도를 가지는 상기 다계층 특징맵의 제2 특징맵을 출력하는 제2 복호화 블록, 상기 제1 특징맵을 입력으로 하여 상기 제1 특징맵보다 높은 해상도를 가지는 제3 특징맵을 출력하는 제3 복호화 블록, 상기 제3 특징맵을 입력으로 가장 큰 해상도를 가지는 상기 다계층 특징맵의 제4 특징맵을 출력하는 제4 복호화 블록을 포함할 수 있다.

[0022] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 제2 복호화 블록, 상기 제3 복호화 블록, 및 상기 제4 복호화 블록은 동일한 구조로 n x n 크기의 커널을 사용하는 합성곱 신경망 층 및 LeakyReLU의 비선형층 포함할 수 있다.

[0023] 본 개시의 다계층 특징맵 복호화 방법, 장치 및 기록매체에 있어서, 상기 제1 복호화 블록, 상기 제3 복호화 블록, 및 상기 제4 복호화 블록의 출력의 해상도는 입력의 해상도보다 높고, 상기 제2 복호화 블록의 출력의 해상도는 입력의 해상도보다 낮을 수 있다.

발명의 효과

[0024] 본 개시는 다계층 특징맵의 각 계층 특징맵을 순차적으로 융합함으로써, 계층 간 중복도를 효과적으로 줄임과 동시에 간결한 신경망 구조로 높은 부호화 성능을 달성할 수 있다.

[0025] 본 개시에 의해 기존 영상 압축 코덱을 이용한 특징맵 압축에 비해 월등히 좋은 부호화 효율을 달성할 수 있으며, 인공 신경망과 영상 압축 코덱을 동시에 사용할 필요 없이 신경망 구조만으로 구성된 특징맵 압축 기술이라는 장점이 있다.

[0026] 또한, 동영상 압축 코덱과 함께 사용하더라도 높은 부호화 효율을 달성할 수 있다.

도면의 간단한 설명

[0027] 도 1은 인공신경망의 하나인 Fast R-CNN을 이용하여 물체 검출 및 분류를 하는 기계작업 결과의 예를 도시한다.

도 2는 다계층 특징맵이 추출되어 기계작업을 수행하는 예를 도시한다.

도 3은 물체 영역분할에 자주 사용되는 인공신경망 모델인 Mask R-CNN의 구조를 도시한다.

도 4는 Mask R-CNN의 FPN을 통해 추출된 다계층 특징맵 P_k의 예를 도시한다.

도 5는 다계층 특징맵의 일 실시예를 도시한다.

도 6은 특징맵 추출 수단은 모바일 장치에 존재하되, 물체분할, 디스패러티맵 추정 또는 영상복원 등과 같은 구체적인 작업 수행 수단은 클라우드 서버에 존재하는 예를 도시한다.

도 7은 신경망 기반 영상의 부호화하는 과정을 도시한다.

도 8은 도 7의 부호화 신경망의 예를 도시한다.

도 9는 도 7의 복호화 신경망의 예를 도시한다.

도 10은 신경망 기반 다계층 특징맵 압축(또는 부/복호화)를 수행하는 과정을 도시한다.

도 11은 도 10의 부호화 신경망의 일 예인 다계층 특징맵 순차적 융합 부호화 신경망을 도시한다.

도 12는 도 10의 복호화 신경망의 일 예인 다계층 특징맵 순차적 융합 복호화 신경망을 도시한다.

도 13은 다계층 특징맵 부호화 방법을 도시한다.

도 14는 다계층 특징맵 복호화 방법을 도시한다.

도 15는 본 개시의 구성 요소를 이용한 다계층 특징맵 부호화 및 복호화 시스템의 상위 수준에서의 제1 실시예를 도시한다.

도 16은 본 개시의 구성 요소를 이용한 다계층 특징맵 부호화 및 복호화 시스템의 상위 수준에서의 제2 실시예를 도시한다.

도 17은 본 실시예에서 사용되는 신경망 블록의 세부 구현을 도시한다.

도 18은 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제1 실시예를 도시한다.

도 19는 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제2 실시예를 도시한다.

도 20은 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제3 실시예를 도시한다.

도 21은 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제4 실시예를 도시한다.

도 22는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제1 실시예를 도시한다

도 23은 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제2 실시예를 도시한다.

도 24는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제3 실시예를 도시한다.

도 25은 본 개시의 다계층 순차적 융합 복호화와 같이 사용될 수 있는 다계층 특징맵 일반 부호화의 예를 도시한다.

도 26는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제4 실시예를 도시한다.

도 27는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제5 실시예를 도시한다.

도 28는 신경망 기반의 융합 잠재표현 엔트로피 부호화의 일 예를 도시한다.

도 29는 이미지에 대한 잠재표현을 채널 방향으로 비균등하게 분할하여 순차적으로 확률 분포를 추정하는 예를 도시한다.

도 30는 잠재표현을 10개의 그룹으로 분할한 후 순차적으로 10번의 확률분포 추정을 수행하는 예를 도시한다.

도 31은 본 개시의 다계층 특징맵 순차적 융합 부/복호화기와 기존 영상 압축 코덱의 부/복호화기에 기반하여 특징맵 압축(또는 부/복호화)를 수행하는 예 및 특징맵 압축(또는 부/복호화)를 위해 인공신경망을 사용하는 예를 도시한다.

도 32는 도 31의 NN video encoder 및 NN video decoder의 세부 내용을 도시한다.

도 33은 본 개시의 다계층 특징맵 순차적 융합 부/복호화기와 기존 영상 압축 코덱을 이용한 융합 잠재표현 부호화 및 복호화의 수행 과정을 도시한다.

도 34는 융합 잠재 표현 특징맵의 구성을 도시한다.

도 35는 특징맵을 기존 영상 압축 코덱에 입력하기 위한 공간적, 시간적, 시공간적 배열을 통한 프레임 패킹을 도시한다.

도 36은 시간적으로 이전에 복호화 된 다계층 특징맵을 이용하여 다계층 특징맵 부/복호화기에 적용하는 블록도의 일 예를 도시한다.

도 37은 다계층 특징맵 순차적 융합 부호화 신경망의 일 예인 도 18에 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 부호화 수단을 영상에 적용하는 신경망의 예를 도시한다.

도 38은 다계층 시간적 문맥을 활용하여 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기에 적용하는 신경망의 예를 도시한다.

도 39은 다계층 시간적 문맥 순차적 융합 수단인 다계층 시간적 문맥 순차적 융합 신경망의 일 예를 도시한다.

도 40은 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 동영상에 적용하는 신경망의 일 예를 도시한다.

도 41은 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 영상에 적용하는 신경망의 일 예를 도시한다.

도 42는 움직임 정보를 사용하지 않고 이전에 복호화 된 각 계층 특징맵의 시간적 문맥을 추출하는 신경망의 일 예를 도시한다.

도 43은 움직임 정보를 사용하여 이전에 복호화 된 각 계층 특징맵의 시간적 문맥을 추출하는 신경망의 일 예를 도시한다.

도 44는 본 개시의 신경망 블록 및 함수 블록도의 세부 구현을 도시한다.

도 45는 움직임 정보를 추출하는 수단인 움직임 정보 추출 신경망의 일 예를 도시한다.

도 46은 움직임 벡터 부호화 수단인 움직임 벡터 부호화 신경망의 일 예를 도시한다.

도 47은 움직임 벡터 복호화 수단인 움직임 벡터 복호화 신경망의 일 예를 도시한다.

도 48은 움직임 보상 수단인 움직임 보상 신경망의 일 예를 도시한다.

도 49는 변형 가능한 합성곱 신경망 층을 사용하여 이전에 복호화 된 각 계측 특징맵의 시간적 문맥을 추출하는 신경망의 일 예를 도시한다.

도 50은 변형 가능한 합성곱 신경망의 일 예를 도시한다.

도 51은 변형 가능한 합성곱 신경망에 쓰이는 오프셋을 추출하는 오프셋 추출 수단 신경망의 일 예를 도시한다.

도 52는 오프셋 부호화 수단인 오프셋 부호화 신경망의 일 예를 도시한다.

도 53은 오프셋 복호화 수단인 오프셋 복호화 신경망의 일 예를 도시한다.

도 54는 변형 가능한 합성곱 신경망 층을 사용한 움직임 보상 수단인 움직임 보상 신경망의 일 예를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0028] 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.

[0029] 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.

[0030] 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.

[0031] 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.

[0032] 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.

[0033] 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.

[0034] 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.

[0035] 물체 분류(Object Classification), 물체 인식(Object Recognition), 물체 검출(Object Detection), 물체 영역

분할(Object Segmentation), 물체 추적(Object Tracking) 등 다양한 기계비전(Machine Vision) 작업이나 해상도 향상(Super-Resolution), 프레임 보간(Frame-Interpolation)등 다양한 영상처리(Image Processing) 작업과 같은 기계작업(Machine Task)에서 인공신경망(ANN: Artificial Neural Network)을 활용하는 사례가 점점 늘어나고 있다.

[0036] 도 1은 인공신경망의 하나인 Fast R-CNN을 이용하여 물체 검출 및 분류를 하는 기계작업 결과의 예를 도시한다.

[0037] 도 1을 참조하면, 도 1의 인식된 물체들은 각각 자동차, 배, 사람, 교통신호 등으로 분류될 수 있고, 각각의 영역은 사각형으로 구성될 수 있으며, 해당 영역이 분류된 물체와 일치하는지에 대한 확률 정보(또는 정확도)가 포함될 수 있다. 여기서, 사각형은 영역의 형태의 일 예이고, 영역의 형태는 원, 다각형 등을 포함할 수 있다. 또한, 영역 내 영역분할을 통해 보다 영역과 물체의 범위가 최대한 일치하도록 할 수 있다.

[0038] 기계작업(Machine Task)을 수행하는 인공신경망 모델은 통상 입력 데이터 혹은 입력 영상으로부터 특징을 추출하는 특징맵 추출 수단과 추출된 특징에 기반하여 구체적인 기계작업을 실제로 수행하는 작업 수행 수단으로 구성될 수 있다. 이때, 입력이 영상 형태인 경우, '추출된 특징'은 통상 특징맵(Feature map)이라 불릴 수 있다. 본 개시에서는 특징맵이라는 표현을 사용하여 발명의 설명을 기재하지만, 맵 형태가 아닌 특징의 경우에도 본 개시를 동일하게 적용할 수 있다.

[0039] 도 2는 다계층 특징맵이 추출되어 기계작업을 수행하는 예를 도시한다.

[0040] 본 개시는 피라미드 구조 특징맵 또는 다 계층 특징맵에 적용 가능할 수 있다. 다계층 특징맵은 서로 다른 해상도를 갖는 특징맵이 여러 계층을 이루는 피라미드 구조를 가질 수 있다. 일 예로, 특징맵이 높은 계층에 속할수록(또는 계층 인덱스가 증가할 수록) 특징맵의 해상도는 감소하며, 낮은 계층에 속할 수록(또는 계층 인덱스가 감소할 수록) 특징맵의 해상도는 증가할 수 있다. 다른 일 예로, 특징맵이 높은 계층에 속할 수록(또는 계층 인덱스가 증가할 수록) 특징맵의 해상도는 증가하며, 낮은 계층에 속할 수록(또는 계층 인덱스가 감소할 수록) 특징맵의 해상도는 감소할 수 있다. 또한, 추후 설명할 도 34와 같이 같은 계층 내의 특징맵끼리는 같은 해상도를 가질 수 있다.

[0041] 다계층 특징맵 내 각 계층의 계층 정보(예, 계층 인덱스, 계층 해상도 등)는 특징맵 정보에 포함되어 기계 작업을 수행함에 있어 전달될 수 있다.

[0042] 본 개시는 이러한 다계층 특징맵에 대한 압축(또는 부/복호화) 방법, 장치 및 기록매체에 관한 것일 수 있다.

[0043] 도 3은 물체 영역분할에 자주 사용되는 인공신경망 모델인 Mask R-CNN의 구조를 도시한다.

[0044] 도 3의 Mask R-CNN 구조에서는, 특징 피라미드 네트워크(FPN: Feature Pyramid Network)가 다계층 특징맵 추출 수단으로 사용되고, 영역 제안 네트워크(RPN: Region Proposal Network)와 관심영역 헤드(ROI Heads: Region Of Interest Heads)가 기계 작업 수행 수단으로 사용될 수 있다.

[0045] 특징 파라미드 네트워크(FPN)는 다계층 특징맵을 추출하는 일 예이며, FPN에 의해 C계층 특징맵과 P계층 특징맵이 추출될 수 있다. 여기서, C계층 특징맵과 P계층 특징맵은 모두 본 개시의 다계층 특징맵일 수 있다.

[0046] 이하, 본 개시에 있어서, 도 3의 P계층 특징맵을 이용하여 발명을 설명한다. 다만, 이는 설명의 편의를 위한 것일 뿐, 도 3의 C계층 특징맵이나, 다른 형태의 다계층 특징맵에 대해서도 본 개시의 설명이 동일하게 적용될 수 있다.

[0047] 입력이 영상인 경우, 특징맵의 형태는 가로x세로 2차원 배열로 나타낼 수 있으며, 하나의 계층의 특징맵은 추후 설명할 도 34와 같이 통상 여러 개의 채널로 구성되어 있으므로, 각 계층의 특징맵은 가로_크기 x 세로_크기 x 채널_개수 만큼의 크기를 갖는 3차원 배열로 표현될 수 있다.

[0048] 즉, 계층 k의 특징맵을 F_k라 하면 F_k는 추출된 특징값들로 구성된 3차원 배열 $F_k[x][y][c]$로 나타낼 수 있다. 이때 x와 y는 각각 특징값의 가로 위치와 세로 위치를 나타내며, c는 채널 인덱스를 나타낼 수 있다. 예를 들어, FPN에서 추출된 다계층 특징맵 C_k 또는 다계층 특징맵 P_k들이 본 개시의 다계층 특징맵 F_k라 할 수 있으며, 3차원 배열 $F_k[x][y][c]$로 표현될 수 있다.

[0049] 도 4는 Mask R-CNN의 FPN을 통해 추출된 다계층 특징맵 P_k의 예를 도시한다.

[0050] 도 4를 참조하면, 도 3의 Mask R-CNN의 FPN을 통해 추출된 다계층 특징맵인 P계층 특징맵 P_k에서 첫 번째 채널

의 예를 확인할 수 있다. 이때 계층 인덱스인 k는 2,3,4,5,6 일 수 있다.

[0051] 다시 말해, FPN에서는 각 계층의 특징맵이 256개의 채널로 구성될 수 있는데, 도 4에서는 각 계층의 특징맵 중 첫번째 채널에 해당하는 특징맵 만을 예시로 나타낸 것일 수 있다. 참고로 FPN에서는 계층이 깊어질수록(계층 인덱스가 증가할 수록) 특징맵의 가로 및 세로의 크기가 입력 영상의 크기(해상도)보다 점점 더 작아질 수 있다.

[0052] 도 5는 다계층 특징맵의 일 실시예를 도시한다.

[0053] 도 5를 참조하면, YOLO v3와 같은 기계 작업을 위한 모델에서도 P 계층 특징맵과 유사한 방법으로 도 5와 같이 세 개의 계층으로 이루어진 다계층 특징맵(Output 1, Output2, Output 3)을 추출하여 이를 기계 작업 수행에 사용할 수 있다.

[0054] 구체적으로, 위의 다계층 특징맵은 YOLO v3 내의 다계층 특징맵 추출 수단으로 darknet 53과 함께 FPN과 유사한 피라미드 구조를 사용하는 경우일 수 있다.

[0055] 기계작업이 대형 서버 뿐만 아니라 모바일 장치를 포함한 다양한 장치에서 광범위하게 사용되면서 특징맵 추출 수단과 작업 수행 수단이 같은 장치내에 존재하지 않고 서로 다른 장치 내에 존재하는 경우도 점점 많아지고 있다.

[0056] 도 6은 특징맵 추출 수단은 모바일 장치에 존재하되, 물체분할, 디스패러티맵 추정 또는 영상복원 등과 같은 구체적인 작업 수행 수단은 클라우드 서버에 존재하는 예를 도시한다.

[0057] 모바일 장치에서 추출된 특징맵이 서버로 전달되며, 전달된 특징맵을 기초로 서버에서 작업을 수행하고, 그에 대한 작업 수행 결과가 다시 모바일 장치로 전달될 수 있다.

[0058] 이러한 예와 같이, 특징맵 추출 수단과 작업 수행 수단이 분리되어 있는 경우에는 추출된 특징맵을 작업 수행 수단에 전달해야만 하며, 전달 또는 저장할 특징맵의 데이터량을 최소화하는 동시에 작업 수행 성능의 저하를 최소화하기 위한 특징맵 부호화 방법이 필요하게 될 수 있다.

[0059] 다른 예로, 특징맵 추출 수단과 작업 수행 수단이 한 장치에 존재하는 경우에라도, 추출된 특징맵을 저장 장치를 통해 저장해 두었다가 이후 작업 수행 수단에서 활용하는 경우가 있을 수 있는데, 이러한 경우에도 상기와 같은 특징맵 압축(또는 부/복호화) 방법이 필요하게 될 수 있다.

[0060] 본 개시에서 영상이라 함은, 카메라를 통해 획득된 자연 영상 뿐 아니라, 컴퓨터 그래픽스, 홀로그램 영상, 신경망을 통해 추출된 특징맵 영상, 초음파 영상 등 다양한 종류의 영상을 의미할 수 있다.

[0061] 도 7은 신경망 기반 영상 압축하는 과정을 도시한다.

[0062] 신경망 기반 영상 압축(또는 부/복호화) 방법의 핵심 구성 요소는, 부호화 신경망, 복호화 신경망, 양자화, 잠재표현 확률 모형, 엔트로피 부호화 및 복호화을 포함할 수 있으며, 영상 압축(또는 부/복호화)를 수행하는 과정을 도 7과 함께 다음 1)~6)와 같이 설명한다.

[0063] 1) 부호화기에서는 부호화하려는 영상(즉, 입력 영상) x를 부호화 신경망을 통해 잠재표현 y로 변환할 수 있다.

[0064] 잠재표현이란, 은닉(잠재)벡터(Latent vector), 잠재 표현(표상)(Latent representation), 및 은닉(잠재)특징맵 (Latent feature map) 중 하나를 의미할 수 있다.

[0065] 2) 잠재표현 y의 각 성분 y_i 은 양자화(Quantization)될 수 있다.

[0066] 3) 양자화된 잠재표현 $\hat{y}$은 학습가능한 잠재표현 확률 모형 $P_{(\hat{y})}(\hat{y})$ 기반의 엔트로피 부호화를 통해 비트스트림(Bitstream) 형태로 복호화기로 전송될 수 있다.

[0067] 4) 부호화기로부터 전송된 비트스트림은 부호화기에서와 동일한 잠재표현 확률 모형 기반의 엔트로피 복호화를 통해 $\hat{y}$으로 복원될 수 있다.

[0068] 5) 복호화 신경망은 $\hat{y}$를 입력으로 하여 영상을 복원할 수 있다. 본 개시에서 복원 영상은 $\hat{x}$와 같이 표기할 수 있다.

[0069] 6) 핵심 구성 요소의 신경망 파라미터는 손실함수와 역전파 알고리즘(Backpropagation algorithm)을 통해 학습

될 수 있다.

[0070] 도 7에서 테두리가 점선으로 표시된 블록은 학습가능한 파라미터 또는 신경망을 의미할 수 있다.

[0071] 본 개시에서 '부호화 블록'이란 신경망 기반의 영상 압축(또는 부/복호화) 방법의 부호화기에서 통상적으로 사용하는 신경망 블록을 의미하며, 다음과 같은 특징을 가질 수 있다.

[0072] 부호화 블록은 입력 영상 또는 다른 부호화 블록의 출력 특징맵을 입력으로 하여 학습된 가중치에 의해 연산된 특징맵을 출력할 수 있다.

[0073] 부호화 블록은 합성곱 신경망 층 (Convolution layer), 전치 합성곱 신경망 층 (Transposed convolution layer), 완전 연결 층 (Fully connected layer), 비전 트랜스포머 (Vision transformer) 등과 같은 신경망 구조를 포함할 수 있다.

[0074] 또한, 부호화 블록은 ReLU (Rectified Linear Unit), Tanh (Tangent hyperbolic), Sigmoid function 등의 비선형 함수를 포함할 수 있다.

[0075] 또한, 부호화 블록은 residual (skip) connection과 같은 element-wise 산술 연산을 포함할 수 있다.

[0076] 부호화 블록은 사용하는 신경망 구조(Convolution layer, Fully connected layer)의 스트라이드 (stride) 또는 은닉 층의 노드(node) 수를 조절하여 입력에 대해 차원 축소 또는 차원 확장을 수행할 수 있으며, 특히 입력의 공간적 해상도(가로 크기, 세로 크기)를 늘리거나 줄일 수 있다.

[0077] 이후 설명에서 부호화 블록(s↑)는 입력의 가로 크기, 세로 크기를 s배 키우는 부호화 블록을 의미하고, 부호화 블록(s↓)는 입력의 가로 크기, 세로 크기를 s배 줄이는 부호화 블록을 의미할 수 있다.

[0078] 본 개시에서 복호화 블록이란 신경망 기반의 영상 압축 방법의 복호화기에서 통상적으로 사용하는 신경망 블록을 의미하며, 다음과 같은 특징을 가질 수 있다.

[0079] 복호화 블록은 입력 영상에 대한 은닉 벡터 또는 다른 복호화 블록의 출력 특징맵을 입력으로 하여 학습된 가중치에 의해 연산된 특징맵을 출력할 수 있다.

[0080] 복호화 블록은 합성곱 신경망 층 (Convolution layer), 전치 합성곱 신경망 층 (Transposed convolution layer), 완전 연결 층 (Fully connected layer), 비전 트랜스포머 (Vision transformer)와 같은 신경망 구조를 포함할 수 있다.

[0081] 또한, 복호화 블록은 ReLU (Rectified Linear Unit), Tanh (Tangent hyperbolic), Sigmoid function 등의 비선형 함수를 포함할 수 있다.

[0082] 또한, 복호화 블록은 residual (skip) connection과 같은 element-wise 산술 연산을 포함할 수 있다.

[0083] 복호화 블록이 사용하는 신경망 구조(Convolution layer, Fully connected layer)의 스트라이드 (stride) 또는 은닉 층의 노드(node) 수를 조절하여 입력에 대해 차원 축소 또는 차원 확장을 수행할 수 있으며, 특히 입력의 공간적 해상도(가로 크기, 세로 크기)를 늘리거나 줄일 수 있다.

[0084] 이후 설명에서 복호화 블록(s↑)는 입력의 가로 크기, 세로 크기를 배 키우는 복호화 블록을 의미하고, 복호화 블록(s↓)는 입력의 가로 크기, 세로 크기를 배 줄이는 복호화 블록을 의미할 수 있다.

[0085] (1) 부호화 신경망과 복호화 신경망

[0086] 신경망 기반의 영상 압축 방법에서는 입력 영상 x을 연속된 부호화 블록을 통해 잠재표현 y로 변환하며, 이러한 부호화 블록의 구성을 부호화 신경망이라 지칭할 수 있다.

[0087] 신경망 기반의 영상 압축 방법에서는 영상에 대한 (양자화된) 잠재표현 $\hat{y}$을 연속된 복호화 블록을 통해 복원 영상 $\hat{x}$으로 변환하며 이러한 복호화 블록의 구성을 복호화 신경망이라 지칭할 수 있다.

[0088] 도 8은 도 7의 부호화 신경망의 예를 도시한다.

[0089] 부호화 신경망은, 부호화할 영상 x를 입력으로 하여 N개의 연속된 부호화 블록을 통해 입력 영상의 잠재표현 y를 출력할 수 있다. 이와 관련하여, 도 8은 N=4인 경우의 실시예를 나타낼 수 있다. 여기서, N은 1이상의 자연수로 2,3,4,5,6,7,8 등을 포함할 수 있다. 또한, N은 기-정의된 값 또는 비트스트림으로부터 시그널링되는 정보

에 의해 결정되는 값일 수 있다.

[0090] 도 9는 도 7의 복호화 신경망의 예를 도시한다.

[0091] 복호화 신경망은, 양자화된 잠재표현 $\hat{y}$을 입력으로 하여 N개의 연속된 복호화 블록을 통해 복원 영상 $\hat{x}$을 출력할 수 있다. 이와 관련하여, 도 9는 N=4 인 경우의 실시예를 나타낼 수 있다. 여기서, N은 1이상의 자연수로 2,3,4,5,6,7,8 등을 포함할 수 있다. 또한, N은 기-정의된 값 또는 비트스트림으로부터 시그널링되는 정보에 의해 결정되는 값일 수 있다.

[0092] 일 예로, 입력 영상의 잠재표현은 채널의 수가 입력 영상과 같거나 입력 영상보다 더 많고 공간적 해상도(높이와 너비)가 더 작아진 특징맵 형식일 수 있다. 다른 예로, 입력 영상의 잠재표현은 채널의 수가 입력 영상과 같거나 입력 영상보다 더 많고 공간적 해상도(높이와 너비)가 더 커진 특징맵 형식일 수 있다.

[0093] 본 개시에서 중간 잠재표현(Intermediate latent representation)이란, 도 8에서 볼 수 있듯이 신경망 기반의 영상 부호화 방법의 부호화 과정에서 사용하는 부호화 블록의 출력을 의미하며, 4개의 부호화 블록을 사용하는 부호화 신경망의 예를 들면, 입력 영상 x를 입력으로 하는 첫 번째 부호화 블록의 출력을 첫 번째 중간 잠재표현 l_1이라고 할 수 있으며, 이전 중간 잠재표현 l_{k-1} (k=2, 3, 4)를 입력으로 하는 k번째 부호화 블록의 출력을 k번째 중간 잠재표현 l_k (k=2, 3, 4)이라 할 수 있다. 이때, l_4는 잠재표현 y와 같을 수 있다. (l_4=y).

[0094] (2) 잠재표현 양자화

[0095] 양자화(Quantization)는 특정 범위 내의 값을 특정 값으로 매핑(Mapping)하는 것으로, 표현할 수 있는 값의 가지 수를 한정할 때 사용할 수 있다.

[0096] 양자화는 크게 균등 양자화와 비균등 양자화로 나눌 수 있는데, 균등 양자화는 하나의 값마다 특정 값으로 매핑되는 범위가 모두 같은 경우이고, 비균등 양자화는 하나의 값마다 매핑되는 범위가 다른 경우를 나타낼 수 있다.

[0097] 신경망 기반 영상 압축 방법에서는 잠재표현의 성분들의 값을 양자화하게 되는데, 부호화 신경망이 비선형 변환을 보장해주기 때문에 균등 양자화를 주로 사용할 수 있다.

[0098] 특히 신경망 기반 영상 압축에서의 양자화는 역전파 알고리즘을 사용할 수 있도록 학습 시점(Training time)에는 양자화 대신 -0.5와 0.5 사이의 균등 노이즈(Uniform noise)를 더하는 과정(수식1)으로 대체할 수 있고, 추론 시점(Inference time)에서는 반올림 함수(수식2)을 사용할 수 있다.

수식1 $$\tilde{\boldsymbol{y}} = \boldsymbol{y} + \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$$

수식2 $$\hat{\boldsymbol{y}} = Round(\boldsymbol{y})$$

[0099]

[0100] 여기서, $\tilde{y}$는 균등 노이즈가 더해진 잠재표현, $\hat{y}$는 양자화된 잠재표현, $\mathcal{U}(a,b)$는 a와 b 사이를 구간으로 하는 균등 분포(Uniform distribution), Round(·)는 통상적인 반올림 연산일 수 있다.

[0101] (3) 잠재표현 확률 모형 $P_{(\hat{y})}(\hat{y})$

[0102] 신경망 기반 영상 압축 방법에서는 양자화된 잠재표현 $\hat{y}$을 엔트로피 부호화하기 위해 $\hat{y}$의 확률분포를 추정하며, 이러한 확률분포를 추정하기 위한 신경망 기반의 확률 모형을 잠재표현 확률 모형 $P_{(\hat{y})}(\hat{y})$이라 한다.

[0103] 잠재표현의 확률 모형 $P_{(\hat{y})}(\hat{y})$은 $\hat{y}$의 모든 성분 $\hat{y}_i$가 서로 독립 확률 변수라는 가정 하에 수식 3과 같이 계산할 수 있다.

[0104] 이때, $P_{(\hat{y})}(\hat{y})$는 $P_{(y)}(y)$와 균등 분포 $\mathcal{U}(-\frac{1}{2},\frac{1}{2})$ 간의 합성곱으로 계산할 수 있기 때문에 최종적으로 잠재표현의 확률 모형은 수식 4와 같이 나타낼 수 있다.

수식 3

$$p_{\hat{\boldsymbol{y}}}(\hat{\boldsymbol{y}}) = \prod_i p_{\hat{y}_i}(\hat{y}_i)$$

[0105]

수식 4

$$p_{\hat{y}_i}(\hat{y}_i) = \left(p_{y_i} * \mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)\right)(\hat{y}_i) = \int_{\hat{y}_i-\frac{1}{2}}^{\hat{y}_i+\frac{1}{2}} p_{y_i}(y)dy$$

[0106]

[0107] 여기서, *는 합성곱, $\mathcal{U}(a,b)$는 구간 (a, b)를 가지는 균등 분포를 의미할 수 있다.

[0108] 양자화되지 않은 잠재표현의 성분 단위 확률 모형 $P_{(y_i)}(y_i)$는 학습 가능한 신경망을 통해 예측할 수 있으며, 본 개시에서는 이러한 과정을 잠재표현 확률 분포 추정이라고 할 수 있다.

[0109] 잠재표현 확률 분포 추정을 수행할 때, p_{y_i}를 가우시안 분포 (μ_i, σ_i)로 가정하여, 분포의 파라미터 (μ_i, σ_i)를 학습가능한 신경망(이를 발명의 핵심 구성에서는 잠재표현 확률 분포 추정 수단이라 칭함)을 통해 예측할 수 있다. 이때 y 또는 $\hat{y}$를 신경망의 추가 입력으로 사용하여 더 높은 예측 성능을 달성할 수 있다.

[0110] (4) 잠재표현의 엔트로피 부호화 및 복호화

[0111] 엔트로피 부호화는 확률변수의 확률분포를 알고 있을 때, 그 확률분포를 이용한 부호화 방법으로, 산술 부호화 (Arithmetic coding) 또는 범위 부호화(Range coding)을 사용할 수 있다.

[0112] 신경망 기반 영상 압축 방법들은 부호화기와 복호화기에서 각각 양자화된 잠재표현 $\hat{y}$에 대해 엔트로피 부호화와 엔트로피 복호화를 수행할 수 있다.

[0113] 이때 엔트로피 부호화 및 복호화는 추정된 잠재표현 확률 분포 $P_{(\hat{y})}(\hat{y})$를 기반으로 수행될 수 있다.

[0114] 신경망 기반 영상 압축은 학습 가능한 파라미터로 구성된 신경망들을 통해 수행되는데, 이러한 신경망의 학습은 학습 데이터로부터 계산된 특정 손실함수를 최소화하는 방향으로 파라미터와 신경망의 가중치들을 갱신하는 역전파 알고리즘 (Backpropagation algorithm)을 통해 수행될 수 있다.

[0115] 영상 압축 모델의 경우, 입력영상과 복원영상 간의 왜곡(Distortion)과 부호화기에서 복호화기로 전송하는 비트스트림의 비트율(Rate)을 동시에 최소화하도록 학습이 이루어질 수 있다.

[0116] 이때, 왜곡 손실함수 L_D는 수식 6과 같이 입력 영상 x와 복원 영상 $\hat{x}$ 간의 평균 제곱 오차 (MSE: Mean Square Error) 또는 MS-SSIM (Multi-Scale Structural Similarity Index Measure)를 이용할 수 있다.

[0117] 수식 6에서 $\mathbf{E}_{\mathbf{x}\sim\mathbf{p}_{\mathbf{x}}}(\cdot)$는 모든 학습데이터의 입력 x 들에 대해 기대값(평균)을 구하는 것을 의미할 수 있다.

[0118] 비트율 손실함수 L_R은 수식 7을 이용하여 계산할 수 있으며, 실제 발생한 잠재표현 분포와 잠재표현 확률 모형을 통해 추정된 잠재표현의 분포 간의 교차 엔트로피(Cross-entropy)라 할 수 있으며, 이런 의미에서 엔트로피 손실함수라 불릴 수 있다.

[0119] 학습을 위한 최종 율-왜곡 손실함수 L_{RD}는 수식 8와 같이 표현할 수 있으며, 왜곡 손실함수 L_D와 엔트로피 손실함수 L_R 간의 비율을 결정하기 위한 상수 λ가 사용되는데, λ에 따라 학습된 모델의 출력에 대해 원하는 화

질 수준 및 비트율을 결정할 수 있다. (일반적으로 λ가 클수록 화질 수준이 높다.)

[0120] x는 학습에 사용할 입력 영상 데이터 샘플, p_x는 학습 데이터의 분포일 수 있다.

수식 6 $L_D = E_{x\sim p_x}(D(x,\hat{x}))$, $D(\cdot,\cdot)$는 MSE 또는 MS-SSIM 등의 왜곡 함수

수식 7 $L_R = E_{x\sim p_x}\left(-\log p_{\hat{y}}(\hat{y})\right)$

수식 8 $L_{RD} = L_R + \lambda \cdot L_D$

[0121]

[0122] 다계층 특징맵은 서로 다른 공간적 해상도를 갖는 여러 개의 특징맵으로 구성되기 때문에, 자연 영상을 부호화할 때와 다른 부호화 방법이 필요할 수 있으며, 본 개시에서는 입력 영상이 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$(은 계층의 수)인 경우의 신경망 기반의 영상 압축 방법을 특히 신경망 기반의 다계층 특징맵 압축 방법으로 기술한다. 일 예로, 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$는 가 작을수록 공간적 해상도가 큰 계층의 특징맵을 의미할 수 있다. 다른 예로, 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$는 가 작을수록 공간적 해상도가 작은 계층의 특징맵을 의미할 수 있다.

[0123] 도 10은 신경망 기반 다계층 특징맵 압축(또는 부/복호화)를 수행하는 과정을 도시한다.

[0124] 신경망 기반의 다계층 특징맵 압축(또는 부/복호화) 방법의 핵심 구성 요소 또한 부호화 신경망, 복호화 신경망, 양자화, 잠재표현 확률 모형, 엔트로피 부호화 및 복호화이며, 신경망 기반의 다계층 특징맵 압축(또는 부/복호화)를 수행하는 과정을 도 10과 1)~6)를 통해 설명한다.

[0125] 1) 부호화기에서는 부호화하려는 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$을 부호화 신경망을 통해 잠재표현 y로 변환할 수 있다.

[0126] 2) 잠재표현 y의 각 성분 y_i을 양자화(Quantization)할 수 있다.

[0127] 3) 양자화된 잠재표현 $\hat{y}$은 학습가능한 잠재표현 확률 모형 $p_{\hat{y}}(\hat{y})$기반의 엔트로피 부호화를 통해 비트스트림(Bitstream) 형태로 복호화기로 전송될 수 있다.

[0128] 4) 부호화기로부터 전송된 비트스트림은 부호화기에서와 동일한 잠재표현 확률 모형 기반의 엔트로피 복호화를 통해 $\hat{y}$으로 복원될 수 있다.

[0129] 5) 복호화 신경망은 $\hat{y}$를 입력으로 하여 다계층 특징맵을 복원할 수 있다. 본 개시에서 복원된 다계층 특징맵은 $\{\hat{F}_k\}_{k=1,\ldots,L}$ 와 같이 표기할 수 있다.

[0130] 6) 핵심 구성 요소의 파라미터는 손실함수와 역전파 알고리즘(Backpropagation algorithm)을 통해 학습될 수 있다.

[0131] 도 10에서 테두리가 점선으로 표시된 블록은 학습가능한 파라미터 또는 신경망을 의미할 수 있다.

[0132] 신경망 기반의 다계층 특징맵 압축은 학습 가능한 파라미터로 구성된 신경망들을 통해 수행되는데, 이러한 신경망을 신경망 기반의 다계층 특징맵 압축 모델이라 하며 이러한 신경망의 학습은 학습 데이터로부터 계산된 특정 손실함수를 최소화하는 방향으로 파라미터와 신경망의 가중치들을 갱신하는 역전파 알고리즘 (Backpropagation algorithm)을 통해 수행될 수 있다.

[0133] 신경망 기반의 다계층 특징맵 압축 모델을 학습할 때, 율-왜곡(Rate-distortion) 최적화 방법 또는 율-성능(Rate-performance) 최적화 방법을 사용할 수 있다.

[0134] 율-왜곡 최적화 방법이란, 입력 다계층 특징맵과 복원된 다계층 특징맵 간의 왜곡(Distortion)과 부호화기에서 복호화기로 전송하는 비트스트림의 비트율(Bit-rate)을 동시에 최소화하도록 학습을 진행하는 것이다.

[0135] 율-성능 최적화 방법이란, 복원된 다계층 특징맵을 통해 수행된 기계 작업 수행 성능(Performance)과 부호화기에서 복호화기로 전송하는 비트스트림의 비트율을 동시에 최소화하도록 학습을 진행하는 것이다.

[0136] 신경망 기반의 다계층 특징맵 압축 모델에서 사용하는 왜곡 손실함수는 수식 9과 같이 입력 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$ 와 복원된 다계층 특징맵 $\{\hat{F}_k\}_{k=1,\ldots,L}$ 간의 평균 제곱 오차 (MSE: Mean Square Error) 또는 MS-SSIM (Multi-Scale Structural Similarity Index Measure)를 각 계층에 따른 가중치 w_k에 대해 가중합을 하여 사용할 수 있다.

[0137] 비트율은 수식 10과 같이 잠재표현 확률 모형에 의해 추정된 잠재표현의 확률 분포와 실제 발생한 잠재표현 간의 교차 엔트로피 (Cross-entropy)로 근사할 수 있다.

[0138] 기계 작업 손실함수 L_P는 압축 및 복원된 다계층 특징맵으로부터 수행된 기계작업의 성능을 의미하며 정답 레이블(Label)과 기계 작업의 추론 결과 사이의 비교를 통해 계산될 수 있다.

[0139] 이때, 기계 작업의 종류에 따라 분류 손실 함수(Classification loss), 바운딩 박스 손실함수 (bounding box loss), 마스크 손실 함수 (mask loss) 등이 사용될 수 있다.

[0140] 율-왜곡 최적화를 위한 손실함수 L_{RD}은 수식 11 와 같이 표현할 수 있으며, 왜곡 손실함수 L_D와 교차 엔트로피 기반 손실함수 L_R 간의 비율을 결정하기 위한 상수 λ가 사용되는데, λ에 따라 학습된 모델의 출력에 대해 원하는 복원 수준 및 비트율을 결정할 수 있다. (일반적으로 λ가 클수록 복원 수준이 높다.)

[0141] 율-성능 최적화를 위한 손실함수 L_{RP}은 수식 12과 같이 표현할 수 있으며, 성능 손실함수 L_P와 교차 엔트로피 기반 손실함수 L_R 간의 비율을 결정하기 위한 상수 λ가 사용되는데, λ에 따라 학습된 모델의 출력에 대해 원하는 복원 수준 및 비트율을 결정할 수 있다. (일반적으로 λ가 클수록 기계 작업의 성능이 높다.)

수식 9 $$L_D = E_{x\sim p_x}\left(\sum_{k=1,\ldots,L} w_k \times D(F_k, \hat{F}_k)\right)$$, $D(\cdot,\cdot)$는 MSE 또는 MS-SSIM 등의 왜곡 함수

수식 10 $$L_R = E_{x\sim p_x}\left(-\log p_{\hat{y}}(\hat{y})\right)$$

수식 11 $$L = L_R + \lambda \times L_D$$

수식 12 $$L = L_R + \lambda \times L_P$$

[0142]

[0143] 다계층 특징맵 부호화를 위해 다계층 특징맵 개를 독립적으로 부호화하거나 개를 하나의 융합특징맵으로 만들어 부호화를 수행할 수 있다. 그러나 전자의 경우는 병렬적으로 번 부호화 함에 따른 복잡도 증가 문제가 있고, 후자의 경우 융합과정과 잠재표현 추출 과정이 직렬적으로 수행됨에 따른 복잡도 증가 문제와 계층간 순차적인 속성을 충분히 활용하지 못한다는 단점이 있다.

[0144] 본 개시는 다계층 특징맵 순차적 융합 부호화 또는 다계층 특징맵 순차적 융합 복호화 과정을 포함하는 신경망 기반의 다계층 특징맵 압축 방법에 관한 것이다. 본 개시의 순차적 융합 부호화 또는 복호화를 통해 다계층 특징맵 융합 과정과 잠재표현 추출과정이 순차적으로 하나의 과정으로 처리됨에 따른 복잡도 감소 및 다계층 특징맵의 순차적 속성 활용을 통한 압축 성능 향상을 꾀할 수 있다.

[0145] 도 11은 도 10의 부호화 신경망의 일 예인 다계층 특징맵 순차적 융합 부호화 신경망을 도시한다.

[0146] 도 11을 참조하면, 다계층 특징맵 순차적 융합 부호화를 수행하기 위한 다계층 특징맵 순차적 융합 부호화 신경망이란, 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$을 효과적으로 부호화하기 위해 연속된 부호화 블록을 통해 융합 잠재표현 f_k을 순차적으로 생성하여 최종적으로 다계층 특징맵에 대한 단일 계층의 융합 잠재표현 y을 출력하는 부호화 신경망일 수 있다.

[0147] 이때 순차적 부호화라 함은, 각 부호화 블록을 거칠 때마다 융합 잠재표현의 계층 인덱스 k가 하나씩 증가하거나 감소하는 것을 의미할 수 있다.

[0148] 특히, 도 10의 구조는 융합 잠재표현 또는 중간 잠재표현의 계층 인덱스 k가 하나씩 증가하는 일 예일 수 있다.

[0149] 융합 잠재표현 f_k이란 다계층 특징맵 부호화를 위한 부호화 신경망에서 이전 부호화 블록의 출력인 중간 잠재표현 l_{k-1} (또는 융합 잠재표현 f_{k-1}) 뿐만 아니라, 그것과 동일한 공간적 해상도를 가진 부호화 대상 특징맵 F_k를 추가 입력으로 하는 부호화 블록의 출력을 의미할 수 있다. 상기 과정에서는 중간 잠재표현과 대상 특징맵이

동일한 해상도를 갖는 경우에 대해서만 설명하였으나, 해상도에 다른 경우에도 해상도를 동일하게 변환하는 과정만 추가하면 상기 과정이 적용될 수 있다. 일 예로, 중간 잠재표현 l_{k-1}과 부호화 대상 특징맵 F_k의 해상도가 다른 경우, 둘 중 하나의 해상도를 변화시켜 중간 잠재표현 l_{k-1}과 부호화 대상 특징맵 F_k의 해상도를 동일하게 하여 상기 과정이 수행될 수 있다.

[0150] 이때, l_{k-1}(또는 f_{k-1})와 F_k를 하나의 부호화 블록의 입력으로 사용하기 위해 채널 간 연결 (concatenation)을 사용할 수 있다.

[0151] 이러한 융합 잠재표현은 다계층 특징맵의 서로 다른 계층에 대한 융합된 정보를 포함할 수 있다.

[0152] 다계층 특징맵 순차적 융합 부호화의 특장점으로는, 모든 계층의 특징맵에 대한 단일 융합 잠재표현으로 부호화하기 때문에 다계층 특징맵의 계층 간 중복도를 최소화하여 부호화 성능을 향상시킬 수 있으며, 모든 계층의 특징맵을 부호화함에도 불구하고, 융합 잠재표현을 매개로 하여 모든 계층의 특징맵의 정보를 융합하기 때문에 기존의 신경망 기반 자연 영상 압축 방법에 비해 큰 복잡도 증가 없이 높은 부호화 성능을 달성할 수 있다.

[0153] 본 개시의 부호화 신경망에 입력되는 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$는 추가적인 전처리 단계를 수행하고 난 후의의 결과 값일 수 있다. 이 경우, 전처리는 해상도에 따른 필터링 과정, 해상도를 변형하는 과정 등을 포함할 수 있다.

[0154] 도 12는 도 10의 복호화 신경망의 일 예인 다계층 특징맵 순차적 융합 복호화 신경망을 도시한다.

[0155] 다계층 특징맵 순차적 융합 복호화를 수행하기 위한 다계층 특징맵 순차적 융합 복호화 신경망이란, 다계층 특징맵을 효과적으로 복호화하기 위해 도 12과 같이 다계층 특징맵에 대한 단일 계층의 융합 잠재표현 y으로부터 연속된 복호화 블록을 통해 다계층 특징맵 $\{\hat{F}_k\}_{k=1,\ldots,L}$을 순차적으로 복호화하는 복호화 신경망일 수 있다.

[0156] 이때, 순차적 복호화라 함은, 각 복호화 블록을 거칠 때마다 복원되는 다계층 특징맵의 계층 인덱스 k가 하나씩 증가하거나 감소하는 것을 의미할 수 있다.

[0157] 특히, 도 11의 구조는 복원되는 다계층 특징맵의 계층 인덱스 k가 하나씩 감소하는 경우의 일 실시예일 수 있다.

[0158] 다계층 특징맵에 대한 단일 융합 잠재표현은 엔트로피 부호화 및 복호화를 위해 양자화되었을 수 있다 (즉, $\hat{y}$).

[0159] 다계층 특징맵 순차적 융합 복호화 신경망은 y를 입력으로 하여 가장 낮은 해상도를 갖는 계층의 특징맵 $\hat{F}_L$부터 가장 높은 해상도를 갖는 계층의 특징맵 $\hat{F}_1$를 순차적으로 복원할 수 있다.

[0160] 이때 k번째 복호화 블록의 출력은 복호화 대상인 $\hat{F}_{L-k+1}$이며, 동시에 다음 복호화 블록의 입력이 되는 것이 특징일 수 있다.

[0161] 복호화 블록을 통해 복호화된 특징맵 $\hat{F}_k$는 그대로 최종 복원 특징맵 $\{\hat{F}_k\}_{k=1,\ldots,L}$이 될 수 있지만, 추가적인 후처리를 거칠 수 있다. 이 경우, 후처리는 해상도에 따른 필터링 과정, 해상도를 변형하는 과정 등을 포함할 수 있다.

[0162] 다계층 특징맵 순차적 융합 복호화의 특장점으로는, 모든 계층의 특징맵을 단일 융합 잠재표현으로부터 복호화하기 때문에 다계층 특징맵의 계층 간 중복도를 최소화하여 부호화 성능을 향상시킬 수 있으며, 모든 계층의 특징맵을 복호화함에도 불구하고, 이전 복호화 블록의 출력을 다음 복호화 블록의 입력으로 사용하는 간결한 구조 때문에 기존의 신경망 기반 자연 영상 압축 방법의 복호화 신경망에 비해 큰 복잡도 증가 없이 높은 부호화 성능을 달성할 수 있다.

[0163] 도 13은 다계층 특징맵 부호화 방법을 도시한다.

[0164] 도 13의 다계층 특징맵 부호화 방법 및 장치는, 입력 영상으로부터 다계층 특징맵을 추출하는 다계층 특징맵 추출 수단[E1], 상기 추출된 다계층 특징맵을 순차적으로 융합 잠재표현으로 변환하여 최종적으로 단일계층의 융합 잠재표현을 출력하는 다계층 특징맵 순차적 융합 부호화 수단[E2], 및 상기 융합 잠재표현을 비트스트림으로

변환하는 융합 잠재표현 부호화 수단[E3]을 포함할 수 있다.

[0165] 여기서, 상기 추출된 다계층 특징맵은 Min-Max 정규화, Z-score 정규화 등의 통상적인 방법으로 정규화 되었을 수 있다[E2-1].

[0166] 융합 잠재표현 부호화 수단은, 영상 압축 코덱을 이용한 융합 잠재표현 부호화[E3-1] 및 신경망 기반의 융합 잠재표현 엔트로피 부호화[E3-2] 중 하나일 수 있다.

[0167] 영상 압축 코덱을 이용한 융합 잠재표현 부호화[E3-1]은, 융합 잠재표현에 대해 양자화, 패딩 또는 크기 조정, 재배열 등의 과정을 수행하여 부호화에 적합한 형식으로 만드는 융합 잠재표현 변환 수단[E3-1-1], 및 변환된 융합 잠재표현을 비트스트림 형태로 부호화하기 위한 영상 압축 코덱의 부호화 수단[E3-1-2]이 포함할 수 있다.

[0168] 신경망 기반의 융합 잠재표현 엔트로피 부호화[E3-2]는, 융합 잠재표현을 양자화하는 융합 잠재표현 양자화 수단[E3-2-1], 양자화된 융합 잠재표현의 확률 분포를 추정하는 융합 잠재표현 확률분포 추정 수단[E3-2-2], 및 추정된 융합 잠재표현의 확률 분포를 이용하여 양자화된 융합 잠재표현을 엔트로피 부호화 하기 위한 융합 잠재표현 엔트로피 부호화 수단[E3-2-3]을 포함할 수 있다.

[0169] 도 13의 [E2]의 다계층 특징맵 순차적 융합 부호화 수단을 동영상에 적용하기 위한 방법으로, 현재 부호화 대상 다계층 특징맵보다 시간적으로 이전에 복호화된 다계층 특징맵을 활용하는 방법[E2-2]을 통해 시간적 문맥을 활용할 수 있다.

[0170] 시간적으로 이전에 복호화된 다계층 특징맵을 활용하는 방법[E2-2]는, 시간적으로 이전에 복호화된 다계층 특징맵을 바탕으로 다계층 시간적 문맥을 추출하는 방법[E2-2-1] 및 시간적으로 이전에 복호화된 다계층 특징맵을 직접 활용하는 방법[E2-2-2] 중 하나일 수 있다.

[0171] 다계층 시간적 문맥을 추출하는 방법[E2-2-1]은, 움직임 정보를 사용하는 다계층 시간적 문맥 추출 수단[E2-2-1-1], 움직임 정보를 사용하지 않는 다계층 시간적 문맥 추출 수단[E2-2-1-2], 및 변형 가능한 합성공 신경망을 사용하는 다계층 시간적 문맥 추출 수단[E2-2-1-3] 중 하나일 수 있다.

[0172] 움직임 정보를 사용하는 다계층 시간적 문맥 추출 수단[E2-2-1-1]는, 이전에 복호화된 다계층 특징맵과 현재 부호화 대상 다계층 특징맵 간의 움직임 정보를 추출하는 움직임 벡터 추출 수단[E2-2-1-1-1], 추출된 움직임 벡터의 잠재표현을 출력하는 움직임 벡터 부호화 수단[E2-2-1-1-2], 움직임 잠재표현을 비트스트림으로 변환하는 움직임 잠재표현 부호화 수단[E2-2-1-1-3], 비트스트림으로부터 움직임 잠재표현을 복호화하는 움직임 잠재표현 복호화 수단[E2-2-1-1-4], 움직임 잠재표현으로부터 움직임 벡터를 복호화하는 움직임 벡터 복호화 수단[E2-2-1-1-5], 및 복호화된 움직임 벡터와 이전에 복호화된 다계층 특징맵을 바탕으로 현재 프레임의 다계층 특징맵을 예측하는 움직임 보상 수단[E2-2-1-1-6]을 포함할 수 있다.

[0173] 변형 가능한 합성공 신경망을 사용하는 다계층 시간적 문맥 추출 수단[E2-2-1-3]은, 이전에 복호화된 다계층 특징맵과 현재 부호화 다계층 특징맵 간을 이용하여 변형 가능한 합성곱 신경망의 오프셋을 추출하는 오프셋 추출 수단[E2-2-1-3-1], 추출된 오프셋의 잠재표현을 출력하는 오프셋화 수단[E2-2-1-3-2], 오프셋 잠재표현을 비트스트림으로 변환하는 오프셋 잠재표현 부호화 수단[E2-2-1-3-3], 비트스트림으로부터 오프셋 잠재표현을 복호화하는 오프셋 잠재표현 복호화 수단[E2-2-1-3-4], 오프셋 잠재표현으로부터 오프셋 벡터를 복호화하는 오프셋 복호화 수단[E2-2-1-3-5], 및 복호화된 오프셋과 이전에 복호화된 다계층을 바탕으로 현재 프레임의 다계층 특징맵을 예측하는 변형 가능한 합성곱 신경망[E2-2-1-3-6]을 포함할 수 있다.

[0174] 신경망 기반의 융합 잠재표현 엔트로피 부호화[E3-2]는, 신경망 기반의 융합 잠재표현 엔트로피 부호화 수단을 동영상에 적용하기 위한 방법으로, 현재 부호화 대상 다계층 특징맵보다 시간적으로 이전에 복호화된 다계층 특징맵을 활용하는 방법[E2-2]을 통해 시간적 문맥을 활용할 수 있다.

[0175] 도 14는 다계층 특징맵 복호화 방법을 도시한다.

[0176] 입력 비트스트림으로부터 다계층 특징맵을 복호화하는 방법 및 장치는, 입력 비트스트림으로부터 잠재표현을 복호화하는 융합 잠재표현 복호화 수단[D1], 및 상기 융합 잠재표현을 입력으로 하여 다계층 특징맵을 순차적으로 복호화하는 다계층 특징맵 순차적 융합 복호화 수단[D2]를 포함할 수 있다.

[0177] 융합 잠재표현 복호화 수단은, 영상 압축 코덱을 이용한 융합 잠재표현 복호화[D1-1], 및 신경망 기반의 융합 잠재표현 엔트로피 복호화[D1-2] 중 적어도 하나일 수 있다.

[0178] 영상 압축 코덱을 이용한 융합 잠재표현 복호화는, 입력 비트스트림으로부터 융합 잠재표현을 복호화하기 위한

영상 압축 코덱의 복호화 수단[D1-1-1], 및 융합 잠재표현에 대해 양자화, 패딩 또는 크기 조정, 재배열 과정의 역과정을 수행하는 융합 잠재표현 역변환 수단[D1-1-2]를 포함할 수 있다.

[0179] 신경망 기반의 융합 잠재표현 엔트로피 부호화는, 양자화된 융합 잠재표현의 확률 분포를 추정하는 융합 잠재표현 확률분포 추정 수단[D1-2-1], 및 상기 추정된 융합 잠재표현의 확률 분포를 이용하여 입력 비트스트림으로부터 양자화된 융합 잠재표현을 엔트로피 복호화하기 위한 융합 잠재표현 엔트로피 복호화 수단[D1-2-2]을 포함할 수 있다.

[0180] 신경망 기반의 융합 잠재표현 엔트로피 복호화 수단[D1-2]를 동영상에 적용하기 위한 방법은, 현재 복호화 대상 다계층 특징맵 보다 시간적으로 이전에 복호화 된 다계층 특징맵을 활용하는 방법[D2-1]을 통해 시간적 문맥을 활용할 수 있다.

[0181] 이전에 복호화 된 다계층 특징맵을 활용하는 방법[D2-1]은, 이전에 복호화된 다계층 특징맵을 바탕으로 다계층 시간적 문맥을 추출하는 방법[D2-1-1], 및 이전에 복호화된 다계층 특징맵을 직접 활용하는 방법[D2-1-2] 중 하나일 수 있다.

[0182] 다계층 시간적 문맥 추출 수단[D2-1-1]은, 움직임 정보를 사용하는 다계층 시간적 문맥 추출 수단[D2-1-1-1], 움직임 정보를 사용하지 않는 다계층 시간적 문맥 추출 수단[D2-1-1-2], 및 변형 가능한 합성공 신경망을 사용하는 다계층 시간적 문맥 추출 수단[D2-1-1-3] 중 하나일 수 있다.

[0183] 움직임 정보를 사용하는 시간적 문맥 추출 수행 방법은, 비트스트림으로 부터 움직임 잠재표현을 복호화하는 움직임 잠재표현 복호화 수단[D2-1-1-1-1], 움직임 잠재표현으로부터 움직임 벡터를 복호화하는 움직임 벡터 복호화 수단[D2-1-1-1-2], 및 움직임 벡터와 이전에 복호화된 다계층 특징맵을 바탕으로 현재 프레임의 다계층 특징맵을 예측하는 움직임 보상 수단[D2-1-1-1-3]을 포함할 수 있다.

[0184] 변형 가능한 합성곱 신경망을 사용하는 시간적 문맥 추출 수행 방법은, 비트스트림으로부터 오프셋 잠재표현을 복호화하는 오프셋 잠재표현 복호화 수단[D2-1-1-3-1], 오프셋 잠재표현으로부터 오프셋 벡터를 복호화하는 오프셋 복호화 수단[D2-1-1-3-2], 및 복호화된 오프셋과 이전에 복호화된 다계층 특징맵을 바탕으로 현재 프레임의 다계층 특징맵을 예측하는 변형 가능한 합성곱 신경망[D2-1-1-3-3]을 포함할 수 있다.

[0185] 도 15는 본 개시의 구성 요소를 이용한 다계층 특징맵 부호화 및 복호화 시스템의 상위 수준에서의 제1 실시예를 도시한다.

[0186] 제1 실시예로 융합 잠재표현 부호화 수단[E3] 및 복호화 수단[D1]으로 영상 압축 코덱을 이용한 융합 잠재표현 부호화[E3-1] 및 복호화[D1-1]를 이용할 수 있다. 각 세부 구성 요소에 대한 설명은 이후 실시예를 통해 보다 자세히 설명한다.

[0187] 도 16은 본 개시의 구성 요소를 이용한 다계층 특징맵 부호화 및 복호화 시스템의 상위 수준에서의 제2 실시예를 도시한다.

[0188] 제2 실시예로 융합 잠재표현 부호화 수단[E3] 및 복호화 수단[D1] 으로 신경망 기반의 융합 잠재표현 엔트로피 부호화[E3-2] 및 복호화[D1-2]를 이용할 수 있다. 각 세부 구성 요소에 대한 설명은 이후 실시예를 통해 보다 자세히 설명한다.

[0189] 신경망을 기반으로 융합 잠재표현 엔트로피 부호화 및 복호화를 수행하는 경우에는 시간적으로 이전에 복호화된 다계층 특징맵을 활용할 수 있다. 시간적으로 이전에 복호화된 다계층 특징맵을 활용하는 방법([E2-2], [D2-1])은 다계층 특징맵 순차적 융합 부호화 수단[E2] 및 복호화 수단[D2]과 신경망 기반의 융합 잠재표현 엔트로피 부호화 수단[E3-2] 및 복호화 수단[D1-2]을 동영상에 적용하기 위한 방법으로 사용할 수 있다.

[0190] 신경망 모델의 부호화 및 복호화 화질 또는 비트율은 앞의 수식 11 또는 수식 12와 같이 매개변수 λ를 사용하여 결정하게 되는데, 이 λ가 특정 임계치보다 낮은 값의 경우에는 시간적으로 이전에 복호화된 다계층 특징맵을 활용하고, 임계치보다 높은 값의 경우에는 시간적으로 이전에 복호화된 다계층 특징맵을 활용하지 않는 스위칭 방법을 사용할 수 있다. 이때, 특정 임계치는 미리 정해 둔 고정된 값을 사용할 수 있고, task의 종류나 입력 특징맵 추출 수단, 혹은 입력 영상이나 특징맵에 따라 적응적으로 결정하여 사용할 수도 있다.

[0191] 도 17은 본 실시예에서 사용되는 신경망 블록의 세부 구현을 도시한다.

[0192] 본 개시의 신경망 블록은, 부호화 신경망에서는 부호화 블록, 복호화 신경망에서는 복호화 블록으로 표현될 수 있다.

[0193] 본 실시예에서는, 다계층 특징맵 $\{F_k\}_{k=1,\ldots,L}$이 특히 P 계층 특징맵 $\{p_k\}_{k=2,\ldots,5}$인 경우를 예로 하여 본 개시의 실시예를 설명한다. 또한, 본 실시예에서 사용되는 신경망 블록의 세부 구현을 도 17를 참조하여 기술할 수 있다.

[0194] Mask R-CNN의 경우 P 계층에서 P_6 계층은 압축하지 않을 수 있다. 이는 P_5 계층과 상관성이 매우 높아서 P_6 대신 P_5 다운샘플링하여 사용해도 기계 작업 성능에는 차이가 없기 때문이다.

[0195] 도 17에서 n X n Conv 블록은 n X n 크기의 커널을 사용하는 합성곱 신경망 층을 의미할 수 있다.

[0196] 도 17에서 n X n Subpixel 블록은 Wenzhe Shi et. al., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network"; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1874-1883 에서 제안된 n X n 크기의 커널을 사용하는 sub-pixel convolution layer를 의미할 수 있다.

[0197] 도 17에서 LeakyReLU는 영상처리를 위한 신경망에서 통상적으로 사용하는 LeakyReLU 비선형 층을 의미할 수 있다.

[0198] 도 17에서 GDN과 IGDN은 J. Balle, et. al., "Density modeling of ´ images using a generalized normalization transformation," arXiv preprint arXiv:1511.06281, 2015.에서 제안된 비선형 층을 의미하며, 이는 신경망 기반의 영상 압축 방법에서 통상적으로 사용될 수 있다.

[0199] 도 17에서 $\hat{p}_H$, $\hat{p}_L$은 다계층 특징맵의 계층들 중 인접한 두 계층의 특징맵을 의미하며, 이때 $\hat{p}_H$는 상대적으로 더 높은 공간적 해상도를 갖는 특징맵, $\hat{p}_L$는 상대적으로 더 낮은 공간적 해상도를 특징맵을 의미할 수 있다.

[0200] 도 17에서 +는 두 특징맵에 대한 성분 간 덧셈(element-wise addition)을 의미하며, Concat은 채널 간 결합(Concatenation)을 의미할 수 있다.

[0201] 도 18은 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제1 실시예를 도시한다.

[0202] 도 18의 제1 실시예는 다계층 특징맵을 순차적으로 융합 잠재표현으로 변환할 때 융합 잠재표현 또는 중간 잠재표현의 계층 인덱스가 순차적으로 증가하는 경우(계층 수 N = 4)의 일 실시예일 수 있다.

[0203] 도 18의 제1 실시예에서는 P계층 특징맵 $\{p_k\}_{k=2,\ldots,5}$를 입력으로 하여 다계층 특징맵을 순차적으로 융합 잠재표현으로 부호화함으로써 최종 융합 잠재표현 y를 출력할 수 있다.

[0204] 도 18의 제1 실시예는 도 13의 [E2]단계의 실시예일 수 있다.

[0205] p_2를 입력으로 하는 부호화 블록1은 Resblock과 Resblock 2↓로 구성되며, 중간 잠재표현인 l_1를 출력할 수 있다.

[0206] p_3과 l_1를 입력으로 하는 부호화 블록2 또한 Resblock과 Resblock 2↓로 구성되며, 융합 잠재표현인 f_2를 출력할 수 있다.

[0207] p_4과 f_2을 입력으로 하는 부호화 블록3 또한 Resblock과 Resblock 2↓로 구성되며, 융합 잠재표현인 f_3를 출력할 수 있다.

[0208] 마지막으로 p_5과 f_3을 입력으로 하는 부호화 블록4는 3X3 합성곱 신경망 층 (2↓과 Attention Module로 구성되며, 최종 융합 잠재표현인 y (=f_4)를 출력할 수 있다.

[0209] Attention Module은 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.에서 제안된 어텐션 모듈을 의미할 수 있다.

[0210] 이때 본 실시예에서의 부호화 블록 외에도 영상처리를 위한 통상의 신경망 기반 부호화 블록들을 조합하여 사용할 수 있다.

[0211] 도 19는 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제2 실시예를

도시한다.

[0212] 도 19의 제2 실시예는 다계층 특징맵을 순차적으로 융합 잠재표현으로 변환할 때 융합 잠재표현 또는 중간 잠재표현의 계층 인덱스가 순차적으로 증가하는 경우(계층수 N = 4)의 일 실시예일 수 있다.

[0213] 도 19를 참조하면, 앞선 실시예들과 유사한 구조를 갖는 다계층 특징맵 순차적 융합 부호화 수단의 추가 실시예가 도시된 것을 확인할 수 있고, P 계층 특징맵 $\{p_k\}_{k=2,\ldots,5}$ 를 입력으로 하여 다계층 특징맵을 순차적으로 융합 잠재표현으로 부호화함으로써 최종 융합 잠재표현 y를 출력할 수 있다.

[0214] p_2 를 입력으로 하는 부호화 블록1은 하나의 5 X 5 합성곱 신경망 층 (2↓과 세 개의 Resblock으로 구성되며, 중간 잠재표현인 l_1 를 출력할 수 있다.

[0215] p_3 과 l_1 를 입력으로 하는 부호화 블록2 또한 하나의 5 X 5 합성곱 신경망 층 (2↓과 세 개의 Resblock, 그리고 하나의 Attention Module로 구성되며, 융합 잠재표현인 f_2 를 출력할 수 있다.

[0216] p_4 과 f_2 을 입력으로 하는 부호화 블록3 또한 하나의 5 X 5 합성곱 신경망 층 (2↓과 세 개의 Resblock으로 구성되며, 융합 잠재표현인 f_3 를 출력할 수 있다.

[0217] 마지막으로 p_5 과 f_3 을 입력으로 하는 부호화 블록4는 5 X 5 합성곱 신경망 층 2↓과 Attention Module로 구성되며, 최종 융합 잠재표현인 y (= f_4)를 출력할 수 있다.

[0218] 도 20은 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제3 실시예를 도시한다.

[0219] 도 20의 제3 실시예는 다계층 특징맵을 순차적으로 융합 잠재표현으로 변환할 때 융합 잠재표현 또는 중간 잠재표현의 계층 인덱스가 순차적으로 증가하는 경우(계층수 N = 3)의 일 실시예일 수 있다.

[0220] 도 20를 참조하면, 앞선 실시예들과 유사한 구조를 갖는 다계층 특징맵 순차적 융합 부호화 수단의 추가 실시예를 확인할 수 있다. YOLO v3 모델에서와 같이 Darknet 53을 특징 추출기로 사용하는 경우를 예로 할 수 있다. 즉 부호화 대상 특징맵 $\{F_k\}_{k=1,\ldots,3}$ 의 계층 수가 3인 경우의 실시예일 수 있다.

[0221] 이러한 경우, 도 20과 같이 세 개의 부호화 블록으로 다계층 특징맵 순차적 융합 부호화 수단을 구성할 수 있다.

[0222] F_1 를 입력으로 하는 부호화 블록1은 하나의 5 X 5 합성곱 신경망 층 (2↓과 세 개의 Resblock, 그리고 하나의 Attention Module로 구성되며, 중간 잠재표현인 l_1 를 출력할 수 있다.

[0223] F_2 와 l_1 를 입력으로 하는 부호화 블록2은 하나의 5 X 5 합성곱 신경망 층 (2↓과 세 개의 Resblock으로 구성되며, 융합 잠재표현인 f_2 를 출력할 수 있다.

[0224] F_3 과 f_2 을 입력으로 하는 부호화 블록3는 5 X 5 합성곱 신경망 층 2↓과 Attention Module로 구성되며, 최종 융합 잠재표현인 y (= f_3)를 출력할 수 있다.

[0225] 도 21은 다계층 특징맵 순차적 융합 부호화 수단인 다계층 특징맵 순차적 융합 부호화 신경망의 제4 실시예를 도시한다.

[0226] 도 21의 제3 실시예는 다계층 특징맵을 순차적으로 융합 잠재표현으로 변환할 때 융합 잠재표현 또는 중간 잠재표현의 계층 인덱스가 순차적으로 증가하는 구조와 순차적으로 감소하는 구조가 병합된 경우의 일 실시예일 수 있다.

[0227] 도 21의 제4 실시예에서는 다계층 특징맵을 순차적으로 융합 잠재표현으로 변환할 때 부호화 블록들의 부호화 순서 및 중간 잠재표현과 융합 잠재표현의 공간적 해상도를 다양하게 구성함으로써 최종 융합 잠재표현의 공간적 해상도를 가변적으로 할 수 있다는 것을 보여줄 수 있다.

[0228] 도 21의 제4 실시예는 도 13의 [E2]단계의 실시예일 수 있다.

[0229] p_2를 입력으로 하는 부호화 블록1은 Relock과 Resblock 2↓로 구성되며, 중간 잠재표현인 l_1를 출력할 수 있다.

[0230] p_3과 l_1를 입력으로 하는 부호화 블록2 또한 Resblock과 Resblock 2↓로 구성되며, 융합 잠재표현인 f_2를 출력할 수 있다.

[0231] p_5를 입력으로 하는 부호화 블록3은 Resblock과 Resblock 2↑로 구성되며, 중간 잠재표현인 l_3를 출력할 수 있다.

[0232] p_4, f_2, l_3를 모두 입력으로 하는 부호화 블록4는 3X3 합성곱 신경망 층 (2↓과 Attention Module로 구성되며, 최종 융합 잠재표현인 y (=f_4)를 출력할 수 있다.

[0233] 본 추가 실시예에서의 최종 융합 잠재표현 y (=f_4)는 제1 실시예에 비해 공간적 해상도가 더 높아지게 될 수 있다.

[0234] 이러한 특징은 부호화하려는 다계층 특징맵의 성격에 따라 최종 융합 잠재표현의 공간적 해상도를 조절하여 부호화 성능을 최적화할 수 있다는 추가적인 장점이 있다.

[0235] 도 22는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제1 실시예를 도시한다.

[0236] 도 22의 제1 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복원되는 다계층 특징맵의 계층 인덱스가 순차적으로 감소하는 경우의 일 실시예일 수 있다.

[0237] 도 22의 제1 실시예에서는 융합 잠재표현 $\hat{y}$를 입력으로 하여 P계층 특징맵을 순차적으로 복호화할 수 있다. 그 결과로 복원된 P계층 특징맵 $\{\hat{p}_k\}_{k=2,\ldots,5}$을 출력할 수 있다.

[0238] 도 22의 제1 실시예는 도 14의 [D2]단계의 실시예 일 수 있다.

[0239] 양자화된 융합 잠재표현 $\hat{y}$를 입력으로 하는 복호화 블록1은 Resblock과 Resblock 2↑로 구성되며, P 계층 특징맵의 특징맵 중 가장 작은 해상도를 갖는 특징맵 $\hat{p}_5$를 출력할 수 있다.

[0240] $\hat{p}_5$를 입력으로 하는 복호화 블록2 또한 Resblock과 Resblock 2↑로 구성되며, P 계층 특징맵 $\hat{p}_4$를 출력할 수 있다.

[0241] $\hat{p}_4$를 입력으로 하는 복호화 블록3 또한 Resblock과 Resblock 2↑로 구성되며, P 계층 특징맵 $\hat{p}_3$을 출력할 수 있다.

[0242] 마지막으로 $\hat{p}_3$을 입력으로 하는 복호화 블록4는 3X3 합성곱 신경망 층 (2↑) 과 Attention Module로 구성되며, P 계층 특징맵의 특징맵 중 가장 큰 해상도를 갖는 특징맵 $\hat{p}_2$를 출력할 수 있다.

[0243] Attention Module은 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.에서 제안된 어텐션 모듈을 의미할 수 있다.

[0244] 이때 본 실시예에서의 복호화 블록 외에도 영상처리를 위한 통상의 신경망 기반 복호화 블록들을 조합하여 사용할 수 있다.

[0245] 도 23은 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제2 실시예를 도시한다.

[0246] 도 23의 제2 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복원되는 다계층 특징맵의 계층 인덱스가 순차적으로 증가하는 구조와 순차적으로 감소하는 구조가 병합된 경우의 일 실시예일 수 있다.

[0247] 도 23의 제2 실시예에서는 다계층 특징맵을 순차적으로 복원할 때 복호화 블록들의 복호화 순서를 다양하게 구

성함으로써 부호화하려는 다계층 특징맵의 성격에 맞게 부호화 성능을 최적화할 수 있음을 보여준다.

[0248] 도 23의 제2 실시예는 도 14의 [D2]단계의 실시예 일 수 있다.

[0249] 양자화된 융합 잠재표현 $\hat{y}$를 입력으로 하는 복호화 블록1은 3X3 합성곱 신경망 층 (2↑) 과 Attention Module로 구성되며, P 계층 특징맵 $\hat{p}_4$를 출력할 수 있다.

[0250] $\hat{p}_4$를 입력으로 하는 복호화 블록2은 Resblock과 Resblock 2↓로 구성되며, P 계층 특징맵 $\hat{p}_5$를 출력할 수 있다.

[0251] $\hat{p}_4$를 입력으로 하는 복호화 블록3은 Resblock과 Resblock 2↑로 구성되며, P 계층 특징맵 $\hat{p}_3$를 출력할 수 있다.

[0252] $\hat{p}_3$를 입력으로 하는 복호화 블록4 또한 Resblock과 Resblock 2↑로 구성되며, P 계층 특징맵 $\hat{p}_2$를 출력할 수 있다.

[0253] 도 24는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제3 실시예를 도시한다.

[0254] 도 24의 제3 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복원되는 다계층 특징맵의 계층 인덱스가 순차적으로 증가하는 경우의 일 실시예일 수 있다.

[0255] 도 24의 제3 실시예는 도 14의 [D2]단계의 실시예 일 수 있다.

[0256] 도 24의 제3 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복호화 블록2, 복호화 블록3, 복호화 블록 4의 입력으로 이전 복호화 블록의 출력 외에도, 추가적인 입력을 가질 수 있음을 보여줄 수 있다.

[0257] 양자화된 융합 잠재표현 $\hat{y}$를 입력으로 하는 복호화 블록1은 Resblock, Resblock 2↑, Resblock, Resblock 2↑, Attention Module, Resblock, Resblock 2↑, Resblock, Subpixel2↑ 블록으로 구성되며, P 계층 특징맵 $\hat{p}_2$를 출력할 수 있다.

[0258] 복호화 블록1의 출력인 $\hat{p}_2$를 입력으로 하는 복호화 블록2는 Feature Mixing 블록으로 구성되며, P 계층 특징맵 $\hat{p}_3$을 출력할 수 있다.

[0259] 복호화 블록2는 융합 잠재표현 $\hat{y}$으로부터 Resblock, Resblock 2↑, Resblock, Resblock 2↑, Attention Module, Resblock, Resblock 2↑으로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0260] 복호화 블록2의 출력인 $\hat{p}_3$를 입력으로 하는 복호화 블록3는 Feature Mixing 블록으로 구성되며, P 계층 특징맵 $\hat{p}_4$을 출력할 수 있다.

[0261] 복호화 블록3는 융합 잠재표현 $\hat{y}$으로부터 Resblock, Resblock 2↑, Resblock, Resblock 2↑으로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0262] 복호화 블록3의 출력인 $\hat{p}_4$를 입력으로 하는 복호화 블록4는 Feature Mixing 블록으로 구성되며, P 계층 특징맵 $\hat{p}_5$을 출력할 수 있다.

[0263] 복호화 블록4는 융합 잠재표현 $\hat{y}$으로부터 Resblock, Resblock 2↑으로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0264] 도 25은 본 개시의 다계층 순차적 융합 복호화와 같이 사용될 수 있는 다계층 특징맵 일반 부호화의 예를 도시한다.

[0265] 본 개시의 다계층 특징맵 순차적 융합 부호화 및 복호화는 반드시 함께 사용되어야 하는 것은 아닐 수 있다. 일예로, 본 개시의 순차적 융합 복호화는 도 25와 같은 다계층 특징맵 일반 부호화와 함께 사용될 수 있다. 다른

예로, 본 개시의 순차적 융합 부호화는 다계층 특징맵 일반 복호화와 함께 사용될 수 있다.

[0266] 다계층 특징맵 부호화를 위해 다계층 특징맵 순차적 융합 부호화 방법을 사용했더라도, 반드시 다계층 특징맵 순차적 융합 복호화를 통해 다계층 특징맵을 복원해야 하는 것은 아니며, 다계층 특징맵을 복원하기 위해 다계층 특징맵 일반 복호화 방법을 이용할 수 있다.

[0267] 마찬가지로, 다계층 특징맵 부호화를 위해 도 25(다계층 특징맵 일반 부호화 방법)와 같이 또 다른 부호화 블록의 조합을 이용했더라도, 그 최종 출력이 단일 계층 특징맵 형식으로 표현될 수 있다면, 그로부터 다계층 특징맵 순차적 융합 복호화를 통해 다계층 특징맵을 복원할 수도 있다.

[0268] 도 26는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제4 실시예를 도시한다.

[0269] 도 26의 제4 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복원되는 다계층 특징맵의 계층 인덱스가 순차적으로 증가하는 경우의 일 실시예일 수 있다.

[0270] 도 26의 제4 실시예는 도 14의 [D2]단계의 실시예 일 수 있다.

[0271] 도 26의 제4 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복호화 블록2, 복호화 블록3, 복호화 블록 4의 입력으로 이전 복호화 블록의 출력 외에도, 추가적인 입력을 가질 수 있음을 보여줄 수 있다.

[0272] 양자화된 융합 잠재표현 $\hat{y}$를 입력으로 하는 복호화 블록1은 5x5 Tconv 2↑, Resblock, Attention Module, 1x1 Conv의 블록의 조합들로 구성되며, P 계층 특징맵 $\hat{p}_2$를 출력할 수 있다.

[0273] 복호화 블록1의 출력인 $\hat{p}_2$를 입력으로 하는 복호화 블록2는 Feature Mixing 블록으로 구성되며, P 계층 특징맵 $\hat{p}_3$을 출력할 수 있다.

[0274] 복호화 블록2는 융합 잠재표현 $\hat{y}$으로부터 5x5 Tconv 2↑, Resblock, Attention Module, 1x1 Conv의 블록의 조합들로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0275] 복호화 블록2의 출력인 $\hat{p}_3$를 입력으로 하는 복호화 블록3는 Feature Mixing 블록으로 구성되며, P 계층 특징맵 $\hat{p}_4$을 출력할 수 있다.

[0276] 복호화 블록3는 융합 잠재표현 $\hat{y}$으로부터 5x5 Tconv 2↑, Resblock, 1x1 Conv의 블록의 조합들로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0277] 복호화 블록3의 출력인 $\hat{p}_4$를 입력으로 하는 복호화 블록4는 Feature Mixing 블록으로 구성되며, P 계층 특징맵 $\hat{p}_5$을 출력할 수 있다.

[0278] 복호화 블록4는 융합 잠재표현 $\hat{y}$으로부터 5x5 Tconv 2↑, Resblock, 1x1 Conv의 블록의 조합들로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0279] TConv 블록은 통상의 전치 합성곱 층(Transposed convolution layer)를 의미할 수 있다.

[0280] 도 27는 다계층 특징맵 순차적 융합 복호화 수단인 다계층 특징맵 순차적 융합 복호화 신경망의 제5 실시예를 도시한다.

[0281] 도 27의 제5 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복원되는 다계층 특징맵의 계층 인덱스가 순차적으로 증가하는 경우의 일 실시예일 수 있다.

[0282] 도 27의 제5 실시예는 도 14의 [D2]단계의 실시예 일 수 있다.

[0283] 도 27의 제5 실시예는 다계층 특징맵을 순차적으로 복호화할 때 복호화 블록2, 복호화 블록3의 입력으로 이전 복호화 블록의 출력 외에도, 추가적인 입력을 가질 수 있음을 보여줄 수 있다. (계층 수 N = 3)

[0284] 양자화된 융합 잠재표현 $\hat{y}$를 입력으로 하는 복호화 블록1은 5x5 Tconv 2↑, Resblock, Attention Module, 1x1 Conv의 블록의 조합들로 구성되며, 다계층 특징맵 $\hat{F}_1$를 출력할 수 있다.

[0285] 복호화 블록1의 출력인 $\hat{F}_1$를 입력으로 하는 복호화 블록2는 Feature Mixing 블록으로 구성되며, 다계층 특징맵 $\hat{F}_2$을 출력할 수 있다.

[0286] 복호화 블록2는 융합 잠재표현 $\hat{y}$으로부터 5x5 Tconv 2↑, Resblock, 1x1 Conv의 블록의 조합들로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0287] 복호화 블록2의 출력인 $\hat{F}_2$를 입력으로 하는 복호화 블록3는 Feature Mixing 블록으로 구성되며, 다계층 특징맵 $\hat{F}_3$을 출력할 수 있다.

[0288] 복호화 블록3는 융합 잠재표현 $\hat{y}$으로부터 5x5 Tconv 2↑, Resblock, 1x1 Conv의 블록의 조합들로 구성된 일련의 신경망 블록을 통해 추출된 특징맵을 추가적인 입력으로 할 수 있다.

[0289] 도 28는 신경망 기반의 융합 잠재표현 엔트로피 부호화의 일 예를 도시한다.

[0290] 도 28을 통해 도 13의 [E3-2] 및 도 14의 [D1-2]의 구체적인 예시를 설명한다.

[0291] [E3-2-1] 관련, 도 28의 Q 블록은 융합 잠재표현 양자화 수단의 일 예로, 융합 잠재표현 y의 모든 성분 y_i에 대해 균등 양자화를 수행하는 것이며(이때 통상의 반올림 연산을 이용할 수 있다), 양자화된 융합 잠재표현 $\hat{y}$를 출력할 수 있다.

[0292] 신경망 기반 영상 압축에서의 양자화는 역전파 알고리즘을 사용할 수 있도록 학습 시점(Training time)에는 양자화 대신 -0.5와 0.5 사이의 균등 노이즈(Uniform noise)를 더하는 과정으로 대체할 수 있고, 추론 시점(Inference time)에서는 반올림 함수를 사용할 수 있다.

[0293] 또한, 학습 과정과 추론 과정의 불일치를 줄이기 위하여 학습 과정에서도 와 같은 반올림 함수를 사용하기도 하는데, 이러한 경우 반올림 함수에 대한 그래디언트(gradient) 계산 시 항등 함수(identity function)을 사용할 수 있다.

[0294] [E3-2-2], [D1-2-1] 관련, 도 28의 엔트로피 모델(Entroy Model)은 융합 잠재표현 확률분포 추정 수단의 일 예로, David Minnen et. al., "Joint autoregressive and hierarchical priors for learned image compression," Advances in neural information processing systems, vol. 31, 2018. 에서 제안된 엔트로피 모델 (entropy model)을 기반으로 구성할 수 있다.

[0295] 융합 잠재표현 $\hat{y}$으로부터 추가 정보(side information)인 hyperprior $\hat{z}$를 먼저 부호화하여 복호화기에 전송함으로써 융합 잠재표현에 대한 더 정확한 확률분포 추정이 가능하게 할 수 있다.

[0296] 또한, 이미 복호화된 융합 잠재표현의 성분들 $\hat{y}_{<i}$를 기반으로 현재 부호화하려는 융합 잠재표현의 성분 $\hat{y}_i$에 대한 추가 정보(context)를 추출하여 이를 부호화에 이용할 수 있다.

[0297] 이외에도 아래와 같은 신경망 기반 기술들에서 사용한 엔트로피 모델(entropy model)을 융합 잠재표현 확률분포 추정 수단으로 사용할 수 있다.

[0298] [E3-2-3], [D1-2-2] 관련, 도 28의 AE와 AD는 각각 융합 잠재표현 엔트로피 부호화 수단과 융합 잠재표현 엔트로피 복호화 수단의 일 예로, 산술 부호화(Arithmetic Encoding) 및 산술 복호화(Arithmetic Decoding)을 나타내며, 산술 부호화 외에도 범위 부호화(Range coding)을 사용할 수 있다.

[0299] 도 29는 이미지에 대한 잠재표현을 채널 방향으로 비균등하게 분할하여 순차적으로 확률 분포를 추정하는 예를 도시한다.

[0300] 도 30는 잠재표현을 10개의 그룹으로 분할한 후 순차적으로 10번의 확률분포 추정을 수행하는 예를 도시한다.

[0301] [E3-2-2], [D1-2-1] 융합 잠재표현 확률분포 추정 수단으로 Space-Channel ConTeXt (SCCTX) entropy model을 사용할 수 있다.

[0302] SCCTX는 도 30(a)와 같이 이미지에 대한 잠재표현을 채널 방향으로 비균등하게 5개의 그룹($\hat{y}^{(1)}$, ... , $\hat{y}^{(5)}$)으

로 분할한 후, 각 잠재표현 그룹 $\hat{y}^{(k)}$에 대해 순차적으로 확률 분포를 추정할 때 이미 복호화된 잠재표현 그룹 $\hat{y}^{(<k)}$의 정보를 추가로 활용하는 방법을 포함할 수 있다. 이러한 추가 정보를 channel context라고 할 수 있다.

[0303] 추가적으로, 도 29과 같이 각 잠재표현 그룹 $\hat{y}^{(k)}$을 공간적으로 checkerboard 패턴으로 분할하여 두 개의 하위 그룹 $\hat{y}_1^{(k)}$, $\hat{y}_2^{(k)}$으로 나누고, 이미 부호화된 한 하위 그룹을 추가 정보로 활용하여 나머지 하나의 하위 그룹의 확률 분포를 추정할 수 있다. 이러한 추가 정보를 spatial context라고 할 수 있다.

[0304] 최종적으로 SCCTX는 잠재표현을 10개의 그룹으로 분할한 후 순차적으로 총 10번의 확률분포 추정을 수행(도 30(b))함으로써 잠재표현의 엔트로피 부호화 효율을 극대화할 수 있다.

[0305] 본 개시의 융합 잠재표현에 대해서도 이러한 SCCTX를 융합 잠재표현 확률분포 추정 수단으로 사용할 수 있다.

[0306] 도 31은 본 개시의 다계층 특징맵 순차적 융합 부/복호화기와 기존 영상 압축 코덱의 부/복호화기에 기반하여 특징맵 압축(또는 부/복호화)를 수행하는 예 및 특징맵 압축(또는 부/복호화)를 위해 인공신경망을 사용하는 예를 도시한다.

[0307] 도 31에서는, 기계 작업을 위한 특징맵 부호화 방식(Feature Coding for Machines)의 2가지 분류 예를 확인할 수 있다.

[0308] 도 32는 도 31의 NN video encoder 및 NN video decoder의 세부 내용을 도시한다.

[0309] 도 31의 (a)는 HEVC나 VVC와 같은 기존 영상 압축 코덱의 부호화기 및 복호화기에 기반하여 특징맵 부호화를 수행하는 예를 나타내고, (b)는 특징맵 부호화를 위해 인공신경망 코덱의 부호화기 및 복호화기에 기반하여 특징맵 부호화를 수행하는 예를 나타낼 수 있다.

[0310] 기계 작업의 특징맵 부호화에서는 특징맵은 통상 다계층 특징맵일 수 있다. 본 개시가 다계층 특징맵 부호화/복호화 방식이나, 다계층은 단일 계층도 포함하므로 본 개시는 다계층 특징맵에 대해서만 한정하는 것은 아닐 수 있다. 이후 설명에서는 다계층 특징맵 부호화에 대해서만 설명한다.

[0311] 도 31은 도 13의 [E3-1]단계 및 도 14의 [D1-1]단계와 관련될 수 있다.

[0312] 도 31의 (a)의 실시예로 본 개시의 다계층 특징맵 순차적 융합 부호화 및 복호화와 영상 압축 코덱을 결합할 수 있다. 도 31의 (a)에서 Format Conversion 부분을 다계층 특징맵 순차적 융합 부호화 신경망을 사용하는 것이다. 도 10의 특징맵 부호화 방법 측면에서 보면, 도 10의 실시예에서 양자화 부분과 엔트로피 부호화 대신 영상 코덱의 부호화기를 사용하는 것일 수 있다. 이는 도 33과 같이 융합 부호화의 출력인 융합 잠재표현을 영상 코덱의 부호화기로 부호화하여 비트스트림을 생성할 수 있다. 복호과정에서는 엔트로피 복호화 방법과 역양자화 과정 대신 영상 코덱의 복호화기를 사용하는 것일 수 있다.

[0313] 도 33은 본 개시의 다계층 특징맵 순차적 융합 부/복호화기와 기존 영상 압축 코덱을 이용한 융합 잠재표현 부호화 및 복호화의 수행 과정을 도시한다.

[0314] 도 33에는 융합 잠재표현 변환 과정이 추가되어 있음을 확인할 수 있다. 이 과정은 다계층 특징맵 순차적 융합 부호화 신경망의 출력인 도 34와 같은 특징맵의 형태를 가지는데 이 특징맵을 영상 압축 코덱 부호화기에 입력으로 바로 사용할 수 없기 때문이다. 통상 특징맵은 부동소수점 형태의 값으로 표현될 수 있다. 영상 압축 코덱에서는 정수 형태의 값 표현을 필요로 하므로 이러한 변환이 필요할 수 있다. 그리고, 영상 압축 코덱은 단일 프레임 또는 연속된 프레임(시퀀스) 형태로 입력을 받으므로 이러한 변환도 필요할 수 있다. 융합 잠재표현 변환은 이 2가지 과정을 수행하는 역할을 할 수 있다.

[0315] 복호화 과정은 부호화 과정의 역순이므로 설명을 생략한다.

[0316] 도 33에서 융합 잠재표현은 일반적으로 특징맵의 형태로 구성되기 때문에 융합 잠재표현 변환 및 역변환 과정은 특징맵 변환 및 역변환 과정을 예로 이하 설명한다.

[0317] 도 34는 융합 잠재 표현 특징맵의 구성을 도시한다.

[0318] 특징맵 변환 및 역변환 과정은 아래 1) ~ 3)으로 구성될 수 있다.

[0319] 1) 특징맵 채널들의 배열(Arrangement) 수단 및 재배열(Rearrangement) 수단

[0320] 2) 특징값의 양자화 수단 및 역양자화 수단

[0321] 3) 특징맵 변환 정보

[0322] 이하, 구체적인 내용을 검토한다.

[0323] 1) 특징맵 채널들의 배열(Arrangement) 및 재배열(Rearrangement) 수단

[0324] 도 34의 특징맵과 같이, 특징맵은 통상 여러 개의 채널로 구성될 수 있다.

[0325] 이 그림에서 채널의 개수는 N_C(또는 c)로, 채널의 가로 크기는 W_C(또는 x)로, 채널의 세로 크기는 H_C(또는 y)로 표시할 수 있다.

[0326] 여러 개의 채널을 부호화하기 위해서는 이 채널들을 영상 코덱의 부호화기에 입력되는 단위인 프레임(또는 픽쳐) 형태로 변환해야 할 수 있다.

[0327] 특징맵 배열 수단에서는 채널들을 프레임으로 변환할 때 공간적(Spatial) 배열 방식, 시간적(Temporal) 배열 방식, 시공간적(Spatio-Temporal) 배열 방식 중 하나를 사용할 수 있다.

[0328] 이렇게 변환된 특징맵 프레임은 영상 압축 코덱의 부호화기에 의해 압축되고 영상 압축 코덱의 복호화기에 의해 복원될 수 있다. 복원된 특징맵 프레임은 특징맵 재배열 수단에 의해 원래의 채널 구성 형태로 재배열 될 수 있다.

[0329] 도 35는 특징맵을 기존 영상 압축 코덱에 입력하기 위한 공간적, 시간적, 시공간적 배열을 통한 프레임 패킹을 도시한다.

[0330] 도 35의 (a), (b), (c)는 각각 공간적 배열, 시간적 배열, 시공간적 배열 방식을 나타낼 수 있다.

[0331] 공간적 배열은 그림 (a)에 보인 바와 같이 특징맵 채널들을 가로로 m개 및 세로로 n개씩 마치 타일과 같은 형태로 배열하여 하나의 특징맵 프레임을 구성하는 것을 지칭하며, m과 n은 그 곱이 배열할 채널의 총 개수 N_c와 같아 지도록 설정된다. 공간적 배열된 특징맵 프레임은 인코딩 수단에서 화면내 예측(Intra-Frame Prediction) 방법을 이용하여 부호화 될 수 있다.

[0332] 시간적 배열은 그림 (b)에 보인 바와 같이 특징맵 채널 각각이 하나의 프레임이 되도록 시간적으로 나열하는 것을 지칭하며, 하나의 특징맵에 대한 총 프레임 수는 그 특징맵을 구성하는 총 채널 수 N_c와 같아지도록 설정될 수 있다. 시간적 배열된 특징맵 프레임들은 영상 압축 코덱의 부호화기에서 화면간 예측(Inter-Frame Prediction) 방법을 이용하여 부호화 될 수 있다. 본 개시에서는 이해를 돕기 위해 영상 부호화 표준에서 많이 사용하는 화면간 예측이란 용어를 사용하였으나 기존의 화면간 예측과 구분하기 위해서는 채널간 예측(Inter-Channel Prediction)이라고 부르는 것이 더 적절할 수 있다.

[0333] 시공간적 배열은 그림 (c)에 보인 바와 같이 공간적으로 배열된 프레임을 시간적으로 배열한 것으로, 하나의 프레임을 구성하는 채널의 개수인 m x n에 총 프레임 수를 곱한 값이 배열할 채널의 총 개수 N_c와 같아 지도록 설정될 수 있다.

[0334] 2) 특징값의 양자화 및 역양자화

[0335] 특징값의 양자화는 균등 양자화 또는 비균등 양자화 방식으로 n-bit 정수로 양자화 될 수 있다.

[0336] 수식13와 수식14는 각각 특징값 $F_{Original}$의 n-bit 균등 양자화 과정과 양자화 후 복원된 특징값 $\tilde{F}_{Converted}$의 균등 역양자화 과정을 나타낼 수 있다.

[0337] 양자화 과정에서는 수식13에서와 같이 정수화 과정의 오차를 줄이기 위해 반올림(rounding)이 적용될 수 있으며, n-bit 정수의 범위를 넘는 값들을 Clipping하는 과정이 포함될 수 있다.

[0338] 양자화와 역양자화를 위해서는 양자화된 결과의 비트수 n값과 특징값의 범위 $Range_{MAX}$값을 설정해 주어야 한다. n값은 8 또는 10과 같이 통상의 영상 부호화기에 입력 가능한 값으로 설정할 수도 있고 압축 성능 향상을 위해 4 또는 6과 같이 더 작은 값으로 설정될 수도 있다. $Range_{MAX}$값은 부호화할 특징값들의 최대값 빼기 최소값으로 구할 수도 있고 미리 정해 둔 값을 사용할 수도 있다.

[0339] 상기한 n 값과 $Range_{max}$ 값은 특징값 양자화 관련 파라미터로서 부호화 과정에서 사용한 값을 복호화 과정에서도 똑같이 사용해야만 하므로 다계층 특징맵 부호화기에 미리 정해 놓거나, 비트스트림에 포함하여 다계층 특징맵 복호화기로 전달할 수 있다.

[0340] 특징값의 양자화는 특징맵 배열 이전 또는 이후에 수행 가능할 수 있다.

[0341] 특징값의 역양자화도 특징맵 재배열 이전 또는 이후에 수행이 가능하나 특징맵 배열 이전에 특징값 양자화를 진행했다면 역양자화는 재배열 이후에 진행하고, 특징맵 배열 이후에 특징값 양자화를 진행했다면 역양자화는 재배열 이전에 진행하여 계산 오차를 줄이는 것이 바람직할 수 있다.

(수식 13)

$$F_{Converted} = Round((F_{Original}) \times (2^n - 1) \ / Range_{max} + 2^{n-1}$$

$$IF, F_{Converted} > 2^{n-1}, \quad F_{Converted} = 2^{n-1}$$

$$F_{Converted} < 0, \quad F_{Converted} = 0$$

[0342]

(수식 14)

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

[0343]

[0344] 3) 특징맵 변환 정보

[0345] 특징값 양자화에 사용되는 비트수 n과 특징값 범위 $Range_{max}$ 가 역양자화 수단에 전달되어야 할 수 있다.

[0346] 또한 특징맵 채널들의 배열이 공간적, 시간적, 시공간적 배열 중 어떤 방식으로 이루어졌는지, 또 공간적이나 시공간적 배열이 사용되었다면 하나의 프레임을 구성하는 가로 채널 개수 또는 세로 채널 개수가 몇 개인지 등의 파라미터들을 특징맵 재배열 수단에서 알고 있어야 원래 채널 구성대로 재배열이 가능할 수 있다.

[0347] 따라서 이러한 특징맵 변환 정보들은 미리 정해 두거나 특징맵 인코딩 수단을 통해 부호화 되어 특징맵 복호화 수단에서 복호화 된 후 특징맵 역변환 수단에 전달되도록 하는 것이 바람직할 수 있다.

[0348] 영상 압축 코덱으로는 HEVC (High Efficiency Video Coding) 또는 VVC (Versatile Video Coding)와 같은 기존 동영상 압축 표준의 부호화기 및 복호화기를 사용할 수 있으며, 이 외에도 정지 영상 압축을 위한 코덱의 부호화기 또는 복호화기를 사용할 수 있다.

[0349] 도 36은 시간적으로 이전에 복호화 된 다계층 특징맵을 이용하여 다계층 특징맵 부/복호화기에 적용하는 블록도의 일 예를 도시한다.

[0350] 다계층 특징맵 순차적 융합 부/복호화 수단 및 엔트로피 부/복호화 수단을 동영상에 적용할 경우에는, 현재 부호화 대상 다계층 특징맵보다 시간적으로 이전에 복호화 된 다계층 특징맵을 활용하여 부호화 성능을 높일 수 있다.

[0351] 본 개시에서 현재 부호화 대상 다계층 특징맵은 $\{p_i\}_{i=2,\ldots,5}$ 라고 표시하며, $\{p_i\}_{i=2,\ldots,5}$보다 이전에 부호화 되어 복호화 된 다계층 특징맵은 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 로 표기할 수 있다.

[0352] 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 은 그림의 파란색 화살표로 표시된 바와 같이, 본 개시의 다계층 특징맵 순차적 융합 부호화기, 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기, 다계층 특징맵 순차적 융합 복호화기의 추가 입력으로 활용할 수 있다.

[0353] 이때, 다계층 특징맵 순차적 융합 부호화기, 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기, 다계층 특징맵 순차적 융합 복호화기 세 가지 구성요소 모두에 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 을 추가 입력으로 활용할 수도 있고, 일부 구성요소에만 추가 입력으로 활용할 수 있다.

[0354] 또한, 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 을 입력으로 하여 다계층 시간적 문맥 추출 수단을 통해 다계층 시간적 문맥인 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 를 출력한 후 이 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 를 그림의 빨간색 화살표로 표시된 바와 같이, 본 개시의 다계층 특징맵 순차적 융합 부호화기, 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기, 다계층 특징맵 순차적 융합 복호화기의 추가 입력으로 활용할 수도 있다.

[0355] 이때, 다계층 특징맵 순차적 융합 부호화기, 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기, 다계층 특징맵 순차적 융합 복호화기 세 가지 구성요소 모두에 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 를 추가 입력으로 활용할 수도 있고, 일부 구성요소에만 추가 입력으로 활용할 수 있다.

[0356] 도 37은 다계층 특징맵 순차적 융합 부호화 신경망의 일 예인 도 18에 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 부호화 수단을 영상에 적용하는 신경망의 예를 도시한다.

[0357] 본 개시의 다계층 특징맵 순차적 융합 부호화 수단을 동영상에 적용할 경우에는, 현재 부호화 대상 다계층 특징맵보다 시간적으로 이전에 복호화 된 다계층 특징맵으로부터 추출된 다계층 시간적 문맥을 활용할 수 있다.

[0358] 현재 부호화 대상 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$ 와 다계층 시간적 문맥인 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 를 입력으로 하여 다계층 특징맵을 순차적으로 융합 잠재표현으로 부호화함으로써 최종 융합 잠재표현 y를 출력할 수 있다.

[0359] 부호화 블록1은 p_2 와 p_2 계층의 시간적 문맥인 $\hat{p}_2^{context}$ 의 채널 간 결합된 결과를 입력으로 할 수 있다.

[0360] 부호화 블록2는 p_3 과 p_3 계층의 시간적 문맥인 $\hat{p}_3^{context}$, l_1 의 채널 간 결합된 결과를 입력으로 할 수 있다.

[0361] 부호화 블록3은 p_4 과 p_4 계층의 시간적 문맥인 $\hat{p}_4^{context}$, f_2 의 채널 간 결합된 결과를 입력으로 할 수 있다.

[0362] 마지막으로 부호화 블록4는 p_5 과 p_5 계층의 시간적 문맥인 $\hat{p}_5^{context}$, f_3 의 채널 간 결합된 결과를 입력으로 할 수 있다.

[0363] 도 36에서 설명한 바와 같이, 도 37의 다계층 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 은 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 으로 대체하여 사용될 수 있다.

[0364] 도 38은 다계층 시간적 문맥을 활용하여 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기에 적용하는 신경망의 예를 도시한다.

[0365] 본 개시의 신경망 기반의 융합 잠재표현 엔트로피 부/복호화 수단을 동영상에 적용하는 경우 추가적으로 다계층 시간적 문맥을 활용할 수 있다.

[0366] 구체적으로, 도 38은 신경망 기반의 융합 잠재표현 엔트로피 부/복화기 신경망의 일 예인 도 28에 다계층 시간적 문맥을 활용하여 신경망 기반의 융합 잠재표현 엔트로피 부/복호화기에 적용하는 신경망의 일 예일 수 있다.

[0367] 도 38의 다계층 시간적 문맥 순차적 융합 수단은 다계층 시간적 문맥인 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 을 입력으로 하여 더 정확한 잠재표현 확률 분포 추정을 위한 정보인 융합 시간적 잠재표현 t를 출력할 수 있다. 또한, t는 도 28의 f_{ep} 에 추가적인 입력으로 사용할 수 있다.

[0368] 앞서 도 36에서 설명한 바와 같이, 도 38의 다계층 시간적 문맥인 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 은 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 으로 대체하여 사용될 수 있다.

[0369] 도 39은 다계층 시간적 문맥 순차적 융합 수단인 다계층 시간적 문맥 순차적 융합 신경망의 일 예를 도시한다.

[0370] 도 39를 참조하면, 다계층 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$를 입력으로 하여 다계층 시간적 문맥을 융합 블록들을 통해 순차적으로 융합 시간적 잠재표현으로 융합함으로써 최종적으로 융합 시간적 잠재표현 t을 출력할 수 있다. 이때, 융합 블록은 신경망 블록으로 앞서 기술한 부호화 블록 또는 복호화 블록일 수 있다.

[0371] P2계층의 시간적 문맥인 $\{\hat{p}_2^{context}\}$를 입력으로 하는 융합 블록1은 5 X 5 2↓ 합성곱 신경망 층과 GDN으로 구성할 수 있으며, 중간 잠재표현인 l_1을 출력할 수 있다. P3계층의 시간적 문맥인 $\{\hat{p}_3^{context}\}$와 l_1을 입력으로 하는 융합 블록2 또한 5 X 5 2↓ 합성곱 신경망 층과 GDN으로 구성할 수 있으며, 융합 잠재표현인 f_2를 출력할 수 있다. P4계층의 시간적 문맥인 $\{\hat{p}_4^{context}\}$와 f_2을 입력으로 하는 융합 블록3 또한 5 X 5 2↓ 합성곱 신경망 층과 GDN으로 구성할 수 있으며, 융합 잠재표현인 f_3을 출력할 수 있다. 마지막으로 P5계층의 시간적 문맥인 $\{\hat{p}_5^{context}\}$와 f_3을 입력으로 하는 융합 블록4는 5 X 5 2↓ 합성곱 신경망 층으로 구성할 수 있으며, 최종 융합 시간적 잠재표현인 t(=f_4)를 출력할 수 있다.

[0372] 이때, 본 실시예에서의 융합 블록 외에도 영상처리를 위한 통상의 신경망 기반 융합 블록들을 조합하여 사용할 수 있다. 앞서 도 36에서 설명한 바와 같이, 도 39의 다계층 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$은 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$으로 대체하여 사용될 수 있다.

[0373] 도 40은 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 동영상에 적용하는 신경망의 일 예를 도시한다.

[0374] 다계층 특징맵 순차적 융합 복호화 수단을 동영상에 적용할 경우, 다계층 시간적 문맥을 활용할 수 있다. 구체적으로, 도 40은 다계층 특징맵 순차적 융합 복호화 신경망의 일 예인 도 22에 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 동영상에 적용하는 신경망의 일 예를 나타내며, 복호화 된 융합 잠재표현 $\hat{y}$와 다계층 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$을 입력으로 하여 순차적으로 P계층 특징맵을 순차적으로 복호화할 수 있다. 그 결과로 복원된 P 계층 특징맵 $\{\hat{p}_i\}_{i=2,\ldots,5}$을 출력할 수 있다. 이때, 다계층 시간적 문맥을 활용하기 위해 추가적인 복호화 블록(도 40의 복호화 블록 5~8)들을 사용할 수 있다.

[0375] 중간 잠재표현인 l_1과 p_5 계층의 시간적 문맥인 $\{\hat{p}_5^{context}\}$를 입력으로 받는 복호화 블록5는 Resblock으로 구성될 수 있으며, P계층 특징맵 $\hat{p}_5$를 출력할 수 있다. 융합 잠재표현인 f_2와 p_4계층의 시간적 문맥인 $\{\hat{p}_4^{context}\}$를 입력으로 받는 복호화 블록6은 Resblock으로 구성될 수 있으며, P계층 특징맵 $\hat{p}_4$를 출력할 수 있다. 융합 잠재표현인 f_3와 p_3 계층의 시간적 문맥인 $\{\hat{p}_3^{context}\}$를 입력으로 받는 복호화 블록7은 Resblock으로 구성될 수 있으며, P계층 특징맵 $\hat{p}_3$를 출력할 수 있다. 마지막으로 융합 잠재표현인 f_4와 p_2계층의 시간적 문맥인 $\{\hat{p}_2^{context}\}$를 입력으로 받는 복호화 블록8는 Resblock으로 구성될 수 있으며, P계층 특징맵 $\hat{p}_2$를 출력할 수 있다.

[0376] 이때, 본 실시예에서의 복호화 블록 외에도 영상처리를 위한 통상의 신경망 기반 복호화 블록들을 조합하여 사용할 수 있다. 앞서 도 36에서 설명한 바와 같이, 도 40의 다계층 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$은 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$으로 대체하여 사용될 수 있다.

[0377] 도 41은 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 영상에 적용하는 신경망의 일 예를 도시한다.

[0378] 구체적으로, 도 41(a)는 다계층 특징맵 순차적 융합 복호화 신경망의 또 다른 일 예인 도 24에 다계층 시간적 문맥을 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 영상에 적용하는 신경망의 다른 일 예를 나타내며, 융합 잠재표현 $\hat{y}$와 다계층 시간적 문맥을 입력으로하여 다른 복호화 블록의 조합을 통하여 P계층 특징맵을 순차적으로 복호화할 수 있다.

[0379] 중간 잠재표현인 l_1과 p_2 계층의 시간적 문맥인 $\hat{p}_2^{context}$를 입력으로 하는 Feature Mixing block1은 도 41(b-1)과 같이 구성할 수 있으며 $\hat{p}_2$를 출력할 수 있다. 중간 잠재표현인 l_2과 $\hat{p}_2$와 p_3계층의 시간적 문맥인 $\hat{p}_3^{context}$를 입력으로 하는 Feature Mixing block2은 도 41(b-2)과 같이 구성할 수 있으며 $\hat{p}_3$를 출력할 수 있다. 중간 잠재표현인 l_3과 $\hat{p}_3$ 와 p_4 계층의 시간적 문맥인 $\hat{p}_4^{context}$를 입력으로 하는 Feature Mixing block3은 도 41(b-2)과 같이 구성할 수 있으며 $\hat{p}_4$를 출력할 수 있다. 마지막으로, 중간 잠재표현인 l_4와 $\hat{p}_4$와 p_5계층의 시간적 문맥인 $\hat{p}_5^{context}$를 입력으로 하는 Feature Mixing block4은 도 41(b-2)과 같이 구성할 수 있으며 $\hat{p}_5$를 출력할 수 있다.

[0380] 앞서 도 36에서 설명한 바와 같이, 도 41의 다계층 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$은 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$으로 대체하여 사용될 수 있다.

[0381] 도 42는 움직임 정보를 사용하지 않고 이전에 복호화 된 각 계층 특징맵의 시간적 문맥을 추출하는 신경망의 일 예를 도시한다.

[0382] 다계층 시간적 문맥을 추출할 때는 움직임 정보를 사용할 수도 있고, 사용하지 않을 수도 있고, 변형 가능한 합성곱 신경망을 사용할 수 있다.

[0383] 도 42를 참조하면, 이전에 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$를 입력으로 하여 각 계층 특징맵의 시간적 문맥인 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$를 출력할 수 있다. 이때, 본 실시예에서의 시간적 문맥 추출 블록 외에도 영상처리를 위한 통상의 시간적 문맥 추출 블록들을 조합하여 사용할 수 있다.

[0384] 도 43은 움직임 정보를 사용하여 이전에 복호화 된 각 계층 특징맵의 시간적 문맥을 추출하는 신경망의 일 예를 도시한다.

[0385] 도 43을 참조하면, 이전에 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$과 현재 부호화 대상의 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$을 입력으로 하여 다계층 특징맵의 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$를 추출할 수 있다.

[0386] 이전에 복호화 된 다계층 특징맵과 현재 부호화 대상의 다계층 특징맵을 입력으로 하는 움직임 정보 추출 수단은 두 입력을 바탕으로 움직임 벡터를 추출할 수 있다. 움직임 벡터를 입력으로 하는 움직임 벡터 부호화 수단은 움직임 벡터의 잠재표현인 움직임 잠재표현을 출력할 수 있다. 움직임 잠재표현을 입력으로 하는 움직임 잠재표현 부호화 수단은 비트스트림을 출력할 수 있다. 이전에 복호화 된 다계층 특징맵과 움직임 벡터를 입력으로 하는 움직임 보상 수단은 다계층 시간적 문맥을 출력할 수 있다.

[0387] 도 44는 본 개시의 신경망 블록 및 함수 블록도의 세부 구현을 도시한다.

[0388] 도 44(a)는 이후 설명할 움직임 정보 추출 수단의 구성 요소인 움직임 잔차 흐름 블록의 일 예를 나타내며 움직임 벡터와 예측된 P계층 현재 부호화 대상의 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$을 입력으로 하여 움직임 잔차 흐름을 추출할 수 있다. 도 44(a)에서 7X7 Conv 블록은 7X7 크기의 커널을 사용하는 합성곱 신경망 층을 의미할 수 있다. 도 44(a)에서 ReLU는 영상처리를 위해 신경망에서 통상적으로 사용하는 ReLU 비선형 층을 의미할 수 있다. 이때, 본 실시예에서의 움직임 잔차 흐름 생성 블록도 외에도 영상처리를 위한 통상의 신경망 기반 움직임 잔차 흐름 생성 블록도들을 조합하여 사용할 수 있다.

[0389] 도 44(b)는 움직임 보상 블록도의 일 예를 나타내며, 움직임 벡터와 이전에 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$을 입력으로 하여 예측된 P계층을 추출할 수 있다. 도 44(b)에서 쌍선형 와핑(bilinear warping)은 이전에 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$의 좌표를 움직임 벡터를 바탕으로 좌표 변환을 해주는 함수일 수 있다. 이때, 본 실시예에서의 움직임 보상 블록도 외에도 영상처리를 위한 통상의 움직임 보상 함수를 조합하여 사용할 수 있다.

[0390] 도 44(c)는 움직임 벡터 크기 조정 블록도의 일 예를 나타내며, 움직임 벡터를 입력으로 하여 크기가 조정된 움

직임 벡터를 추출할 수 있다. 도 44(c)에서 쌍선형 보간법(bilinear interpolation)은 움직임 벡터의 크기를 조정해주는 함수일 수 있다. 이때, 본 실시예에서의 움직임 벡터 크기 조정 블록도 외에도 영상처리를 위한 통상의 크기 조정 함수를 조합하여 사용할 수 있다.

[0391] 도 45는 움직임 정보를 추출하는 수단인 움직임 정보 추출 신경망의 일 예를 도시한다.

[0392] 도 45를 참조하면, 이전에 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 과 현재 부호화 대상의 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$ 을 입력으로 하여 움직임 정보를 추출할 수 있다.

[0393] 현재 부호화 대상의 P5계층 p_5 과 이전에 복호화 된 P5계층 $\hat{p}_{5,ref}$ 과 0으로 채워진 특징맵을 입력으로 하는 움직임 잔차 흐름 블록1은 P5계층 크기의 중간 움직임 잔차 흐름인 v_1 을 출력한 후 가로와 세로의 크기를 두 배씩 키워 V_1 을 출력할 수 있다. 이전에 복호화 된 P4계층 $\hat{p}_{4,ref}$ 과 V_1 을 입력으로 하는 움직임 보상블록1은 쌍선형 와핑(bilinear warping)을 통해 P4계층 크기의 예측된 P계층 m_1 을 출력할 수 있다. 현재 부호화 대상의 P4계층 p_4 과, V_1 과 m_1 을 입력으로 하는 움직임 잔차 흐름 블록2는 P4 계층 크기의 중간 움직임 잔차 흐름인 v_2 를 생성 후 V_1 과 성분 간 덧셈 후 가로와 세로의 크기를 두 배씩 키워 V_2 를 출력할 수 있다. 이전에 복호화 된 P3계층 $\hat{p}_{3,ref}$ 과 V_2 을 입력으로 하는 움직임 보상블록2은 쌍선형 와핑(bilinear warping)을 통해 P3계층 크기의 예측된 P계층 m_2 을 출력할 수 있다. 현재 부호화 대상의 P3계층 p_3 과, V_2 과 m_2 을 입력으로 하는 움직임 잔차 흐름 블록3은 P3계층 크기의 중간 움직임 잔차 흐름인 v_3 를 생성 후 V_2 과 성분 간 덧셈 후 가로와 세로의 크기를 두 배씩 키워 V_3 를 출력할 수 있다. 이전에 복호화 된 P2계층 $\hat{p}_{2,ref}$ 과 V_3 을 입력으로 하는 움직임 보상블록3은 쌍선형 와핑(bilinear warping)을 통해 P2계층 크기의 예측된 P계층 m_3 을 출력할 수 있다. 현재 부호화 대상의 P2계층 p_2 과, V_3 과 m_3 을 입력으로 하는 움직임 잔차 흐름 블록4는 P2계층 크기의 중간 움직임 잔차 흐름인 v_4 를 생성 후 V_3 과 성분 간 덧셈 후 최종적으로 움직인 벡터 를 출력할 수 있다.

[0394] 도 46은 움직임 벡터 부호화 수단인 움직임 벡터 부호화 신경망의 일 예를 도시한다.

[0395] 도 46을 참조하면, 움직임 벡터 V를 입력으로 하여 움직임 벡터를 움직임 잠재표현으로 부호화함으로써 최종 움직임 잠재표현 y_{MV} 을 출력할 수 있다. 움직임 벡터 V를 입력으로 하는 움직임 벡터 부호화 수단은 4개의 5 X 5 2↓ 합성곱 신경망 층과 3개의 GDN층으로 구성할 수 있으며, 최종 움직임 잠재표현인 y_{MV}을 출력할 수 있다.

[0396] 도 47은 움직임 벡터 복호화 수단인 움직임 벡터 복호화 신경망의 일 예를 도시한다.

[0397] 도 47을 참조하면, 움직임 잠재표현 $\hat{y}_{MV}$ 를 입력으로 하여 움직임 벡터를 복호화함으로써 최종 움직임 벡터 $\hat{V}$ 를 출력할 수 있다. 움직임 잠재표현인 $\hat{y}_{MV}$ 를 입력으로 하는 움직임 벡터 부호화 수단은 4개의 5 X 5 2↓ 합성곱 신경망 층과 3개의 GDN층으로 구성되며, 최종 움직임 벡터인 $\hat{V}$ 을 출력할 수 있다.

[0398] 도 48은 움직임 보상 수단인 움직임 보상 신경망의 일 예를 도시한다.

[0399] 도 48을 참조하면, 움직임 벡터 $\hat{V}$ 과 이전에 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 을 입력으로 하여 움직임 보상을 통해 최종 다계층 시간적 정보 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 를 출력할 수 있다.

[0400] 움직임 벡터 $\hat{V}$ 와 이전에 복호화 된 P2 계층 $\hat{p}_{2,ref}$ 을 입력으로 하는 움직임 보상블록1은 쌍선형 와핑(bilinear warping)을 통해 예측된 현재 프레임의 P2계층인 $P_2^{\text{predicted}}$ 을 출력할 수 있다. 예측된 현재 프레임의 P2계층 $P_2^{\text{predicted}}$ 을 입력으로 하는 시간적 문맥 추출 블록1은 Resblock과 3 X 3 합성곱 신경망 층으로 구성할 수 있으며, 최종 P2계층의 시간적 $\hat{p}_2^{context}$ 를 출력할 수 있다.

[0401] 움직임 벡터 $\hat{V}$ 를 입력으로 하는 움직임 벡터 크기 조정 블록1은 쌍선형 보간법(bilinear interpolation)을 통해 움직임 벡터를 P3계층 크기만큼 줄인 P3계층 크기의 움직임 벡터 $\hat{V}_3$ 을 출력할 수 있다. P3계층 크기의 움직임 벡터 $\hat{V}_3$ 와 이전에 복호화 된 P3계층 $\hat{p}_{3,ref}$ 을 입력으로 하는 움직임 보상블록2은 쌍선형 와핑(bilinear warping)을 통해 예측된 현재 프레임의 P3계층인 $P_3^{\text{predicted}}$ 을 출력할 수 있다. 예측된 현재 프레임의 P3계층 $P_3^{\text{predicted}}$ 을 입력으로 하는 시간적 문맥 추출 블록2은 Resblock과 3 X 3합성곱 신경망 층으로 구성할 수 있으며, 최종 P3계층의 시간적 정보 $\hat{p}_3^{context}$ 를 출력할 수 있다.

[0402] 이때, 본 실시예에서의 시간적 문맥 추출 블록 외에도 영상처리를 위한 통상의 신경망 기반 시간적 문맥 추출 블록들을 조합하여 사용할 수 있다.

[0403] 도 49는 변형 가능한 합성곱 신경망 층을 사용하여 이전에 복호화 된 각 계측 특징맵의 시간적 문맥을 추출하는 신경망의 일 예를 도시한다.

[0404] 도 49를 참조하면, 이전에 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 과 현재 부호화 대상의 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$ 을 입력으로 하여 다계층 특징맵의 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$ 를 추출할 수 있다.

[0405] 이전에 복호화 된 다계층 특징맵과 현재 부호화 대상의 다계층 특징맵을 입력으로 하는 오프셋 추출 수단은 두 입력을 바탕으로 변형 가능한 합성곱 신경망에 쓰일 오프셋을 추출할 수 있다. 오프셋을 입력으로 하는 오프셋 부호화 수단은 오프셋의 잠재표현인 오프셋 잠재표현을 출력할 수 있다. 오프셋 잠재표현을 입력으로 하는 오프셋 잠재표현 부호화 수단은 비트스트림을 출력할 수 있다. 이전에 복호화 된 다계층 특징맵과 오프셋을 입력으로 하는 변형 가능한 합성곱 신경망은 다계층 시간적 문맥을 출력할 수 있다.

[0406] 도 50은 변형 가능한 합성곱 신경망의 일 예를 도시한다.

[0407] 도 50을 참조하면, 단일 계층 특징맵을 입력으로 하는 오프셋 추출 블록을 통해 오프셋을 추출한 후, 해당 오프셋과 입력된 단일 계층 특징맵을 바탕을 입력으로 하는 변형으로 합성곱 신경망 블록을 거쳐 단일 계층 특징맵을 출력할 수 있다.

[0408] 오프셋 추출 블록과 변형 가능한 합성곱 신경망 블록은, 본 실시예에서의 블록 이외에도 영상처리를 위한 통상의 다른 합성곱 신경망 층을 사용할 수 있다.

[0409] 도 51은 변형 가능한 합성곱 신경망에 쓰이는 오프셋을 추출하는 오프셋 추출 수단 신경망의 일 예를 도시한다.

[0410] 도 51을 참조하면, 이전에 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 과 현재 부호화 대상의 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$ 을 입력으로 하여 다계층 오프셋 $\{V_i\}_{i=2,\ldots,5}$ 을 추출할 수 있다.

[0411] 현재 부호화 대상의 다계층 특징맵 $\{p_i\}_{i=2,\ldots,5}$ 과 이전에 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$ 을 입력으로 하는 오프셋 추출 블록은 현재 부호화 대상의 다계층 특징맵과 같은 크기와 채널의 다계층 오프셋을 출력할 수 있다. 이때, 오프셋 추출 블록은 2개의 3 X 3 합성곱 신경망으로 구성할 수 있지만 오프셋 추출 해당 블록은 영상처리를 위한 통상의 다른 합성곱 신경망 층을 조합하여 사용할 수 있다.

[0412] 도 52는 오프셋 부호화 수단인 오프셋 부호화 신경망의 일 예를 도시한다.

[0413] 구체적으로, 도 52는 다계층 특징맵 순차적 융합 부호화 신경망의 일 예인 도 18을 활용하여 다계층 특징맵 순차적 융합 부호화 수단을 오프셋 부호화에 적용한 신경망의 일 예를 나타낼 수 있다.

[0414] 다계층 오프셋 $\{V_i\}_{i=2,\ldots,5}$ 을 입력으로 하여 다계층 오프셋을 순차적으로 융합 잠재표현으로 부호화함으로써 최종 융합 잠재표현 y를 출력할 수 있다. 부호화 블록 1은 V_2 를 입력으로 할 수 있다. 부호화 블록 2은 V_3 과 l_1 의 채널 간 결합된 결과를 입력으로 할 수 있다. 부호화 블록 3은 V_4 과 l_2 의 채널 간 결합된 결과를 입력으로 할 수 있다. 마지막으로 부호화 블록 4은 V_5 과 l_3 의 채널 간 결합된 결과를 입력으로 할 수 있다. 이때, 부호화 블록은

영상처리를 위한 통상의 다른 합성곱 신경망 층을 조합하여 사용할 수 있다.

[0415] 도 53은 오프셋 복호화 수단인 오프셋 복호화 신경망의 일 예를 도시한다.

[0416] 구체적으로, 도 53은 다계층 특징맵 순차적 융합 복호화 신경망의 일 예인 도 22를 활용하여 다계층 특징맵 순차적 융합 복호화 수단을 오프셋 복호화에 적용한 신경망의 일 예를 나타낼 수 있다.

[0417] 복호화된 융합 잠재표현 $\hat{y}$를 입력으로 하여 순차적으로 다계층 오프셋을 순차적으로 복호화한다. 그 결과로 복원된 다계층 오프셋 $\{\hat{V}_i\}_{i=2,\ldots,5}$을 출력할 수 있다. 중간 잠재표현인 l_1을 입력으로 하는 복호화 블록 1은 $\hat{V}_5$을 출력할 수 있다. 중간 잠재표현인 l_2을 입력으로 하는 복호화 블록 2는 $\hat{V}_4$를 출력할 수 있다. 중간 잠재표현인 l_3을 입력으로 하는 복호화 블록 3은 $\hat{V}_3$을 출력할 수 있다. 최종적으로 중간 잠재표현인 l_4을 입력으로 하는 복호화 블록 4는 $\hat{V}_2$을 출력할 수 있다. 이때, 복호화 블록은 영상처리를 위한 통상의 다른 합성곱 신경망 층을 조합하여 사용할 수 있다.

[0418] 도 54는 변형 가능한 합성곱 신경망 층을 사용한 움직임 보상 수단인 움직임 보상 신경망의 일 예를 도시한다.

[0419] 도 54를 참조하면, 다계층 오프셋 $\{\hat{V}_i\}_{i=2,\ldots,5}$와 이전에 복호화 된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$을 입력으로 하여 변형 가능한 합성곱 신경망을 통해 최종 다계층 시간적 정보 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$를 출력할 수 있다.

[0420] 다계층 오프셋$\{\hat{V}_i\}_{i=2,\ldots,5}$와 이전에 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$을 입력으로 하는 변형 가능한 합성곱 신경망 블록 1은 변형 가능한 합성곱 신경망을 통해 $\{f_i\}_{i=2,\ldots,5}$을 출력할 수 있다. $\{f_i\}_{i=2,\ldots,5}$와 이전에 복호화된 다계층 특징맵 $\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$을 입력으로 하는 움직임 보상 블록 2는 3 X 3 합성곱 신경망 층으로 구성할 수 있으며, 예측된 현재 프레임의 다계층 특징 맵 $\{p_i^{predicted}\}_{i=2,\ldots,5}$을 출력할 수 있다. 예측된 현재 프레임의 다계층 특징맵 $\{p_i^{predicted}\}_{i=2,\ldots,5}$을 입력으로 하는 시간적 문맥 추출 블록은 Resblock과 3 X 3합성곱 신경망 층으로 구성할 수 있으며, 최종 다계층의 시간적 문맥 $\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$을 출력할 수 있다. 이때, 본 실시예에서의 신경망 블록 외에도 영상처리를 위한 통상의 신경망 기반 블록들을 조합하여 사용할 수 있다.

[0421] 본 개시의 입력 영상으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵을 추출하는 단계는 다계층 특징맵 추출부에서 수행될 수 있다. 상기 다계층 특징맵을 연속된 부호화 블록들을 통해 순차적으로 부호화하여 단일 계층의 융합 잠재표현을 출력하는 단계는 융합 잠재표현 출력부에서 수행될 수 있다. 상기 단일 계층의 융합 잠재표현을 비트스트림으로 부호화하는 단계는 비트스트림 부호화부에서 수행될 수 있다. 또한, 다계층 특징맵 추출부, 융합 잠재표현 출력부 및 비트스트림 부호화부는 다계층 특징맵 부호화 장치에 포함될 수 있다. 각 부의 구체적인 내용은 위에 수단 또는 방법에서 구체적으로 설명한 바, 이하 생략한다.

[0422] 본 개시의 비트스트림으로부터 계층 형태로 복수의 특징맵들을 포함하는 다계층 특징맵에 대한 단일 계층의 융합 잠재표현을 복호화하는 단계는 비트스트림 복호화부에서 수행될 수 있다. 상기 단일 계층의 융합 잠재표현으로부터 연속된 복호화 블록들을 통해 상기 다계층 특징맵을 순차적으로 복호화하는 단계는 다계층 특징맵 복호화부에서 수행될 수 있다. 비트스트림 복호화부 및 다계층 특징맵 복호화부는 다계층 특징맵 복호화 장치에 포함될 수 있다. 각 부의 구체적인 내용은 위에 수단 또는 방법에서 구체적으로 설명한 바, 이하 생략한다.

[0423] 본 개시의 예시적인 방법들은 설명의 명확성을 위해서 동작의 시리즈로 표현되어 있지만, 이는 단계가 수행되는 순서를 제한하기 위한 것은 아니며, 필요한 경우에는 각각의 단계가 동시에 또는 상이한 순서로 수행될 수도 있다. 본 개시에 따른 방법을 구현하기 위해서, 예시하는 단계에 추가적으로 다른 단계를 포함하거나, 일부의 단계를 제외하고 나머지 단계를 포함하거나, 또는 일부의 단계를 제외하고 추가적인 다른 단계를 포함할 수도 있다.

[0424] 본 개시의 다양한 실시 예는 모든 가능한 조합을 나열한 것이 아니고 본 개시의 대표적인 양상을 설명하기 위한 것이며, 다양한 실시 예에서 설명하는 사항들은 독립적으로 적용되거나 또는 둘 이상의 조합으로 적용될 수도

있다.

[0425] 또한, 본 개시의 다양한 실시 예는 하드웨어, 펌웨어(firmware), 소프트웨어, 또는 그들의 결합 등에 의해 구현될 수 있다. 하드웨어에 의한 구현의 경우, 하나 또는 그 이상의 ASICs(Application Specific Integrated Circuits), DSPs(Digital Signal Processors), DSPDs(Digital Signal Processing Devices), PLDs(Programmable Logic Devices), FPGAs(Field Programmable Gate Arrays), 범용 프로세서(general processor), 컨트롤러, 마이크로 컨트롤러, 마이크로 프로세서 등에 의해 구현될 수 있다.

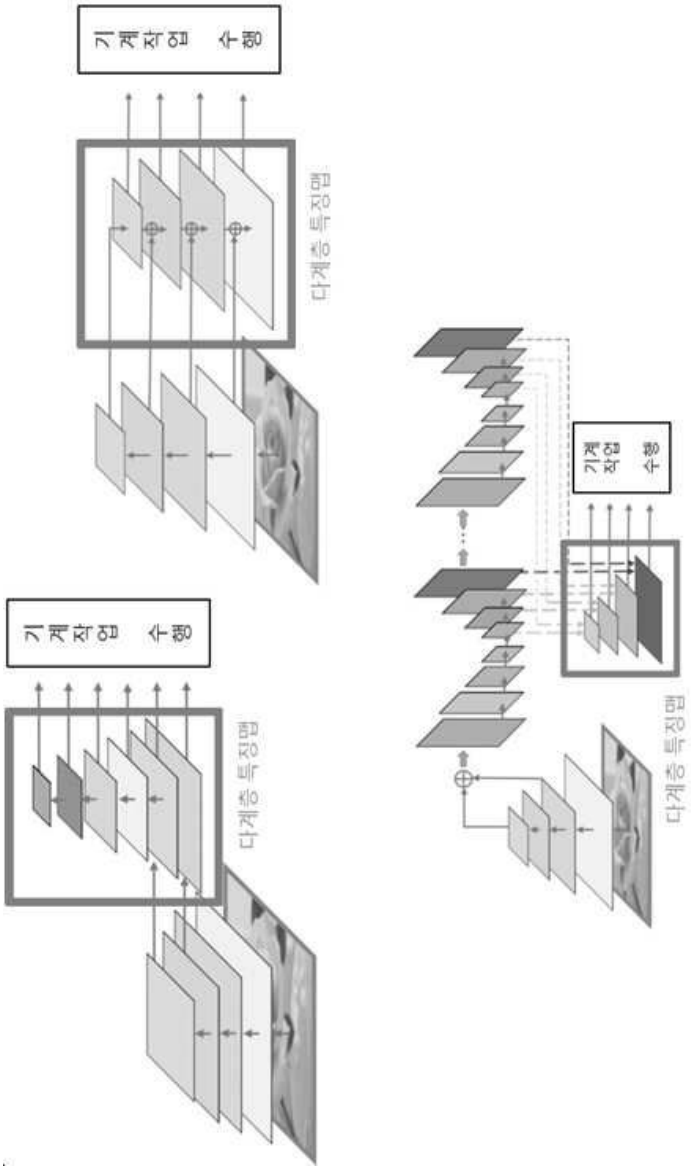
[0426] 본 개시의 범위는 다양한 실시 예의 방법에 따른 동작이 장치 또는 컴퓨터 상에서 실행되도록 하는 소프트웨어 또는 머신-실행 가능한 명령들(예를 들어, 운영체제, 애플리케이션, 펌웨어(firmware), 프로그램 등), 및 이러한 소프트웨어 또는 명령 등이 저장되어 장치 또는 컴퓨터 상에서 실행 가능한 비-일시적 컴퓨터-판독가능한 기록매체(non-transitory computer-readable medium)를 포함한다.

도면

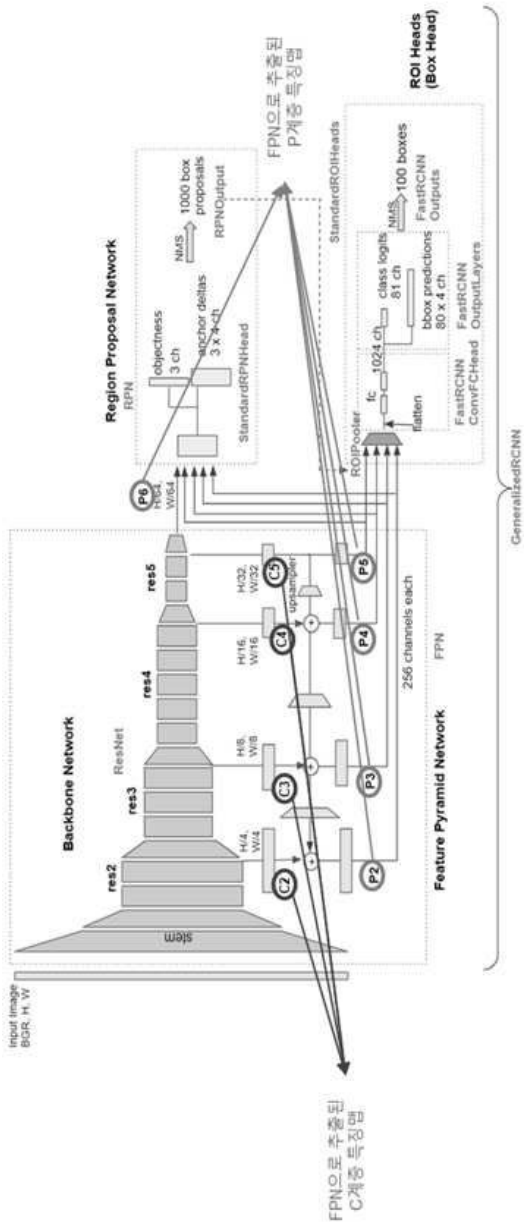
도면1



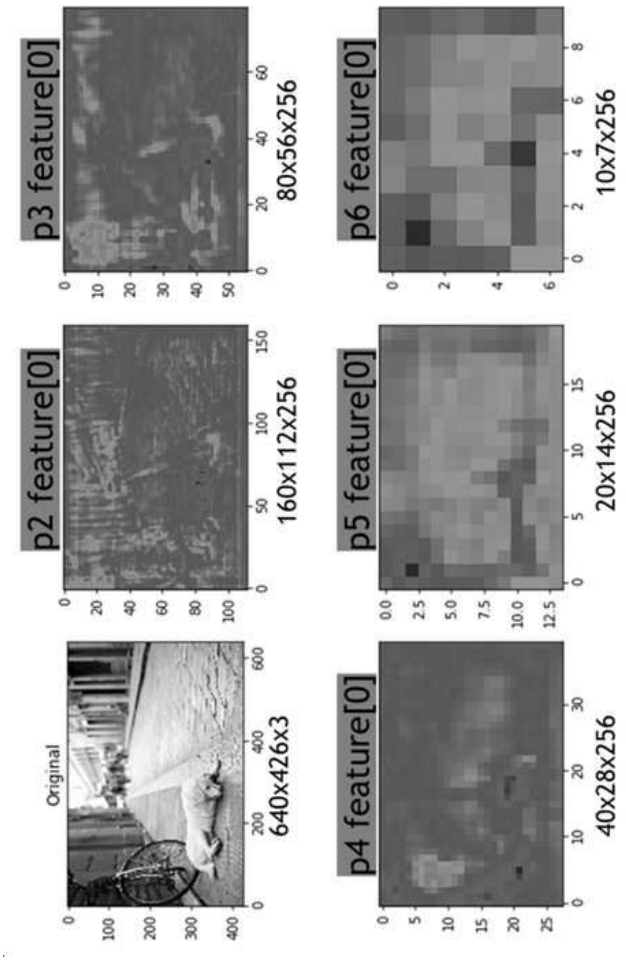
도면2



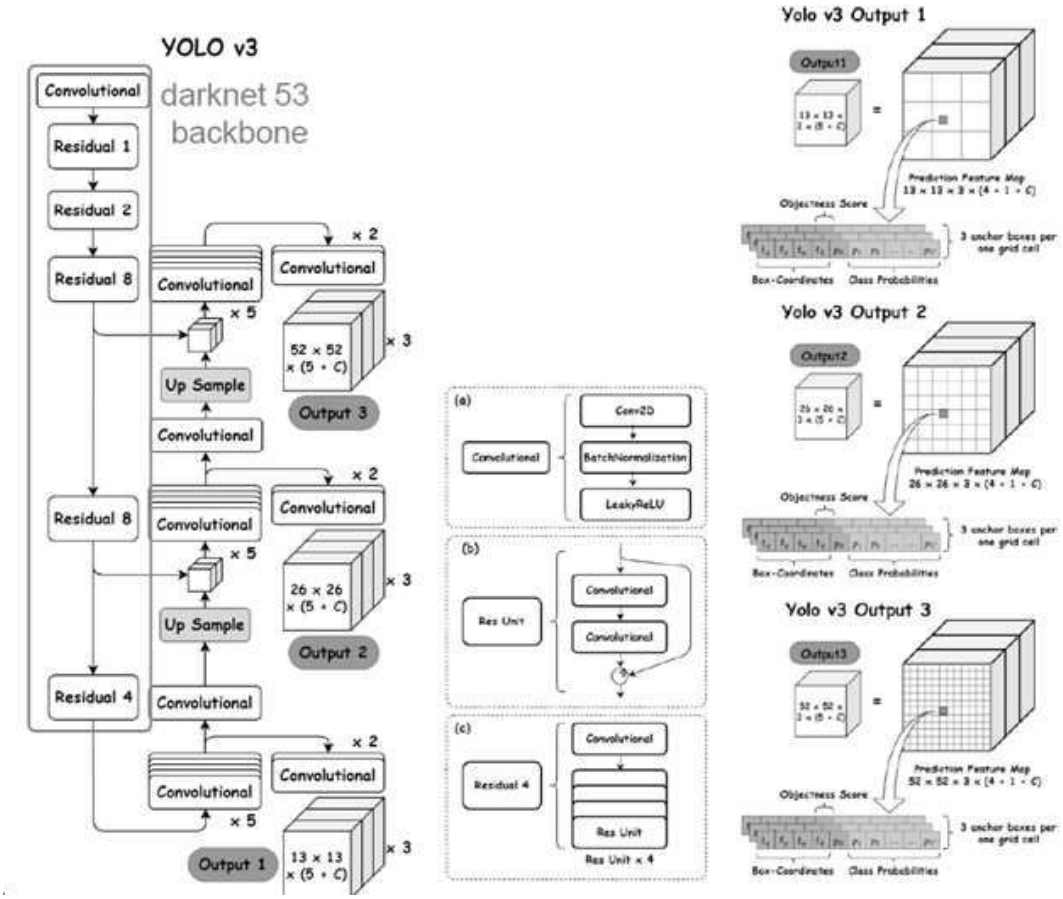
도면3


Backbone Network
ResNet
stem
res2
res3
res4
res5
Feature Pyramid Network
FPN
256 channels each
Region Proposal Network
RPN
StandardRPNHead
objectness
3 ch
anchor deltas
3 x 4 ch
NMS
1000 box proposals
RPNOutput
StandardROIHeads
ROIPooler
fc
1024 ch
flatten
FastRCNN ConvFCHead
FastRCNN OutputLayers
class logits
81 ch
bbox predictions
80 x 4 ch
NMS
100 boxes
FastRCNN Outputs
ROI Heads (Box Head)
GeneralizedRCNN
upsample
input image
BGR, H, W
FPN으로 추출된 P계층 특징맵
FPN으로 추출된 C계층 특징맵

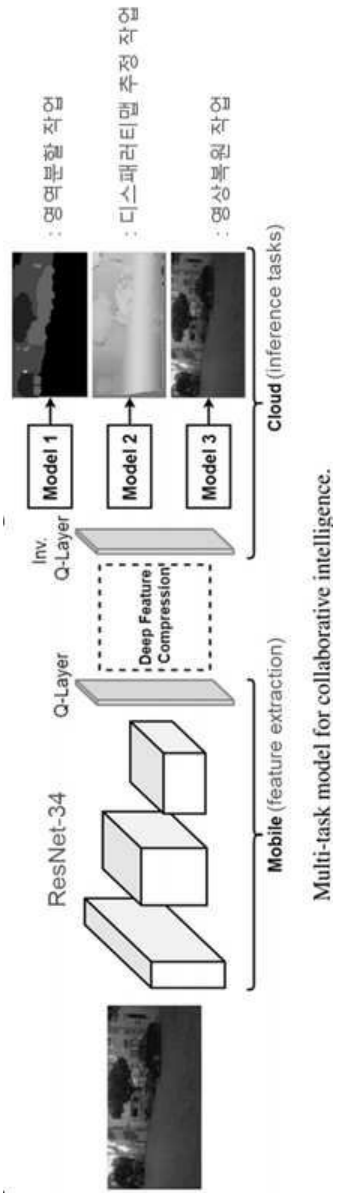
도면4


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

도면5

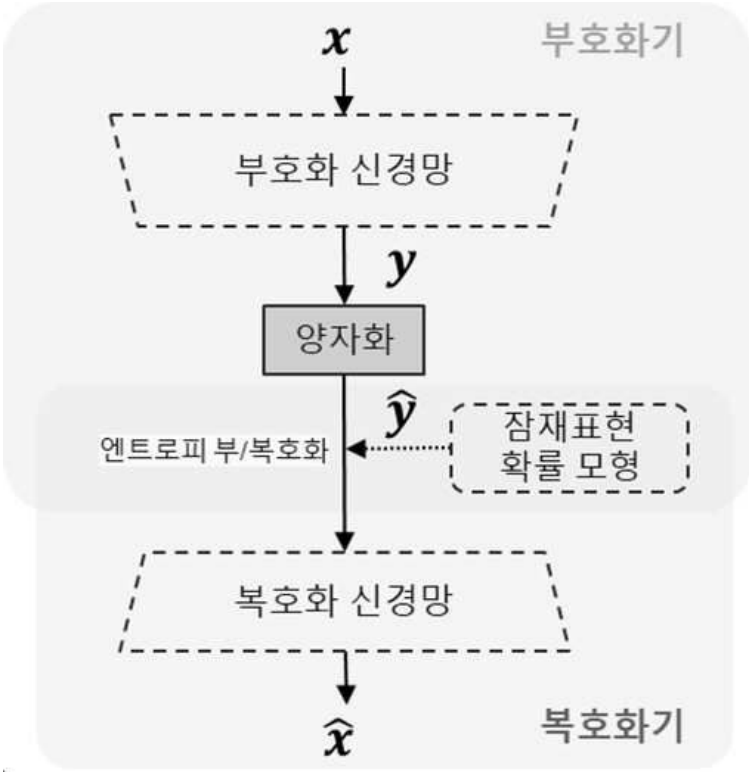

YOLO v3
darknet 53
backbone
Convolutional
Residual 1
Residual 2
Residual 8
Residual 8
Residual 4
Up Sample
Up Sample
Output 1
Output 2
Output 3
52 x 52
26 x 26
13 x 13
Yolo v3 Output 1
Yolo v3 Output 2
Yolo v3 Output 3
Prediction Feature Map
Objectness Score
Box Coordinates
Class Probabilities
3 anchor boxes per one grid cell
Conv2D
BatchNormalization
LeakyReLU
Res Unit
Residual 4

도면6

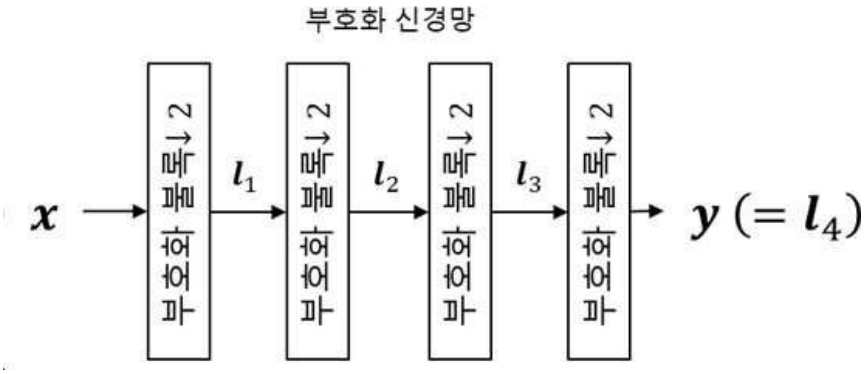


Multi-task model for collaborative intelligence.

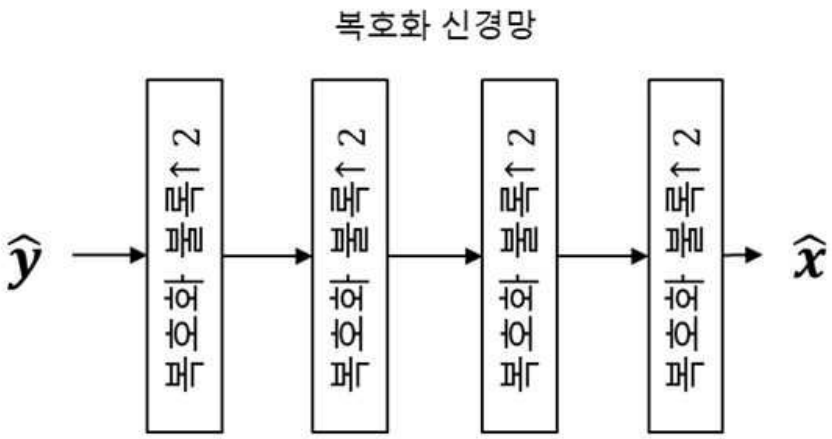
도면7


x
부호화기
부호화 신경망
y
양자화
ŷ
엔트로피 부/복호화
잠재표현
확률 모형
복호화 신경망
x̂
복호화기

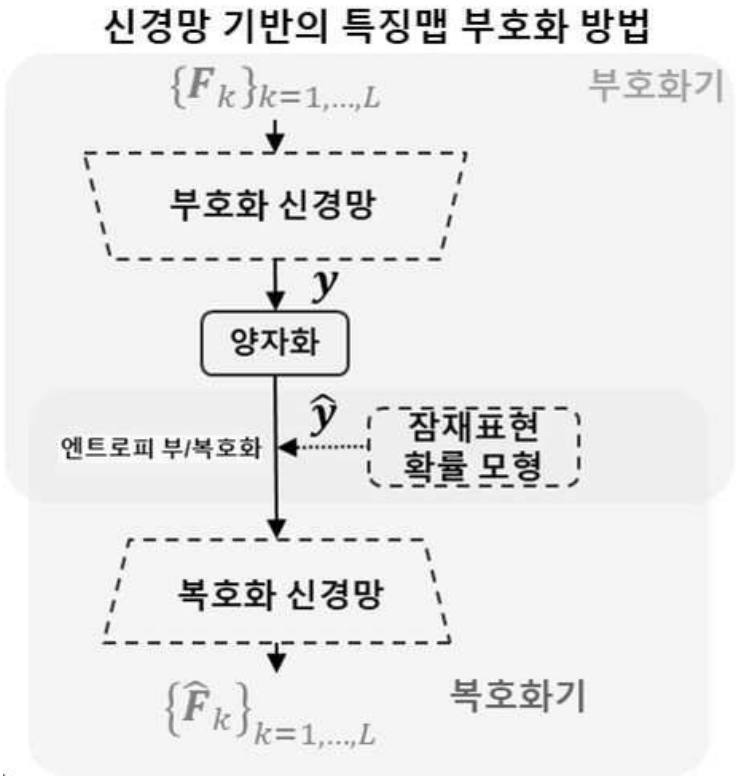
도면8


부호화 신경망
x
합성곱 k=5 s=2
l1
합성곱 k=5 s=2
l2
합성곱 k=5 s=2
l3
합성곱 k=5 s=2
y (= l4)

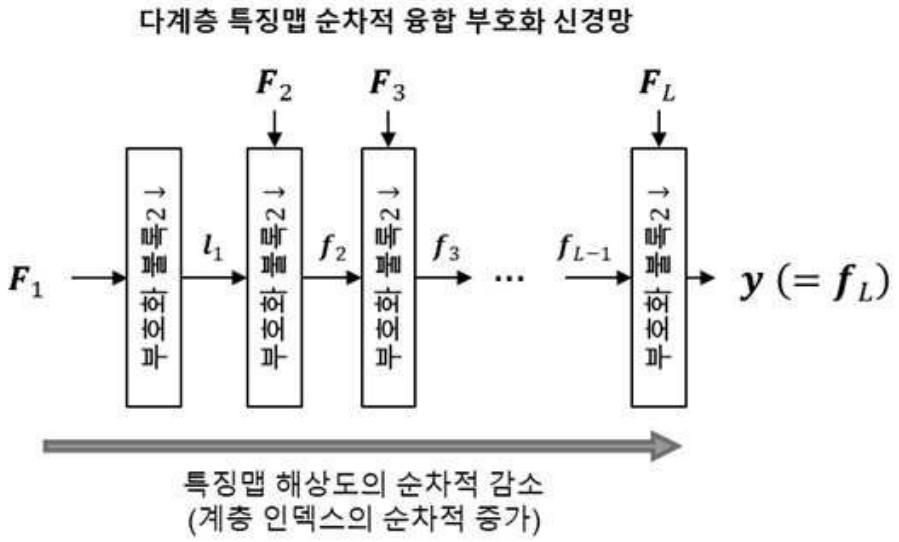
도면9


복호화 신경망
$\hat{y}$
2↑콘볼루션 버퍼
2↑콘볼루션 버퍼
2↑콘볼루션 버퍼
2↑콘볼루션 버퍼
$\hat{x}$

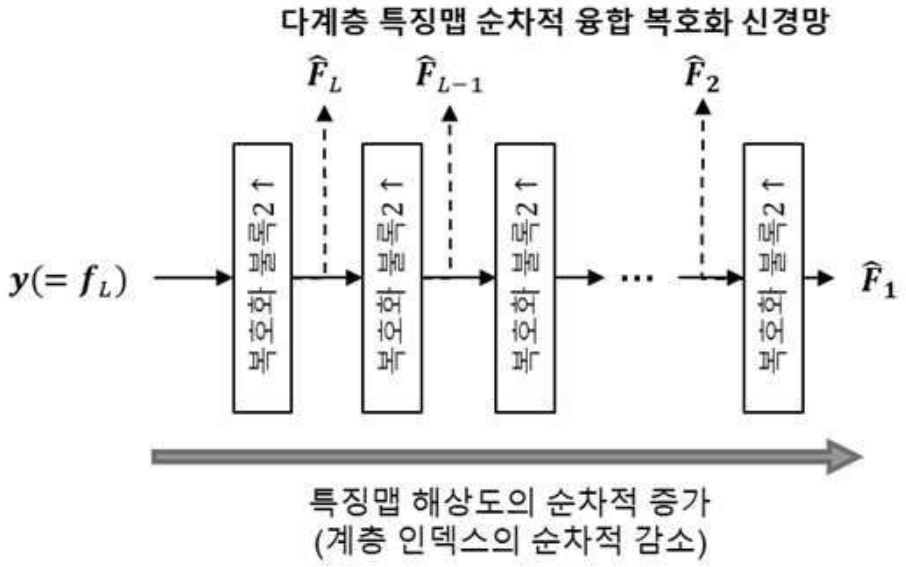
도면10


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

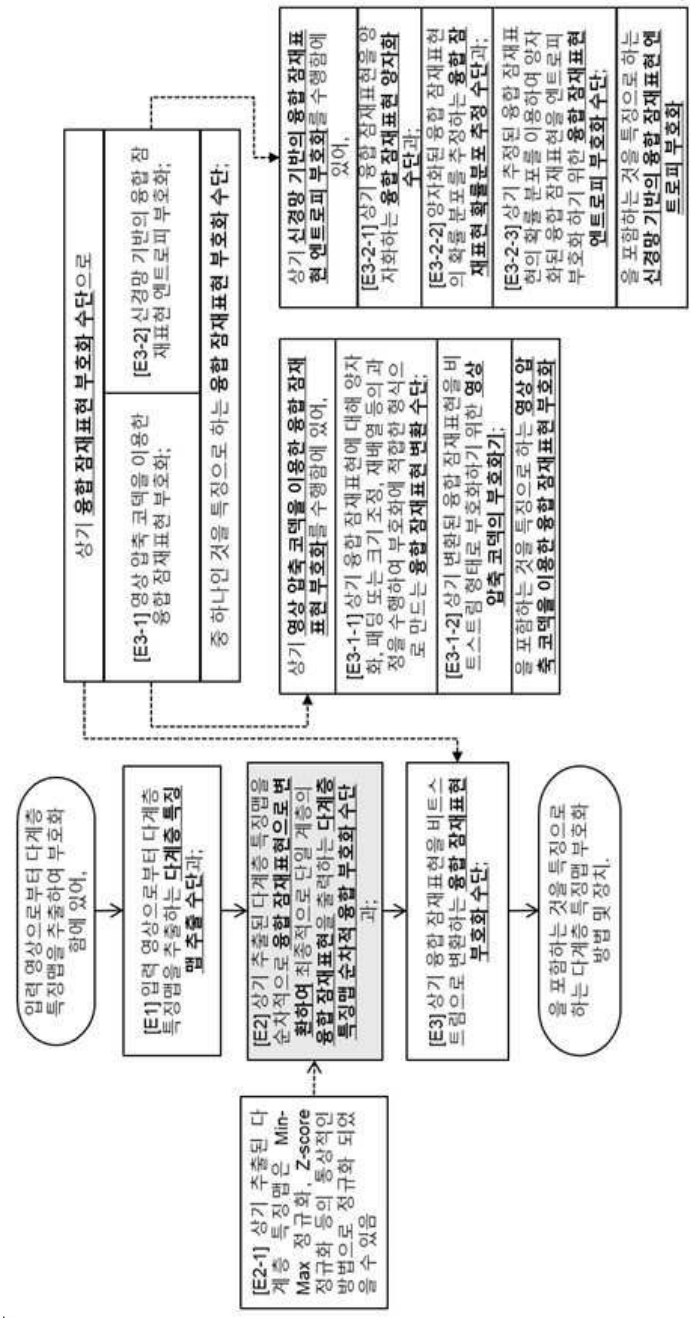
도면11


다계층 특징맵 순차적 융합 부호화 신경망
F_1
F_2
F_3
F_L
부호화 블록 2↓
l_1
f_2
f_3
f_{L-1}
$y\ (=f_L)$
특징맵 해상도의 순차적 감소
(계층 인덱스의 순차적 증가)

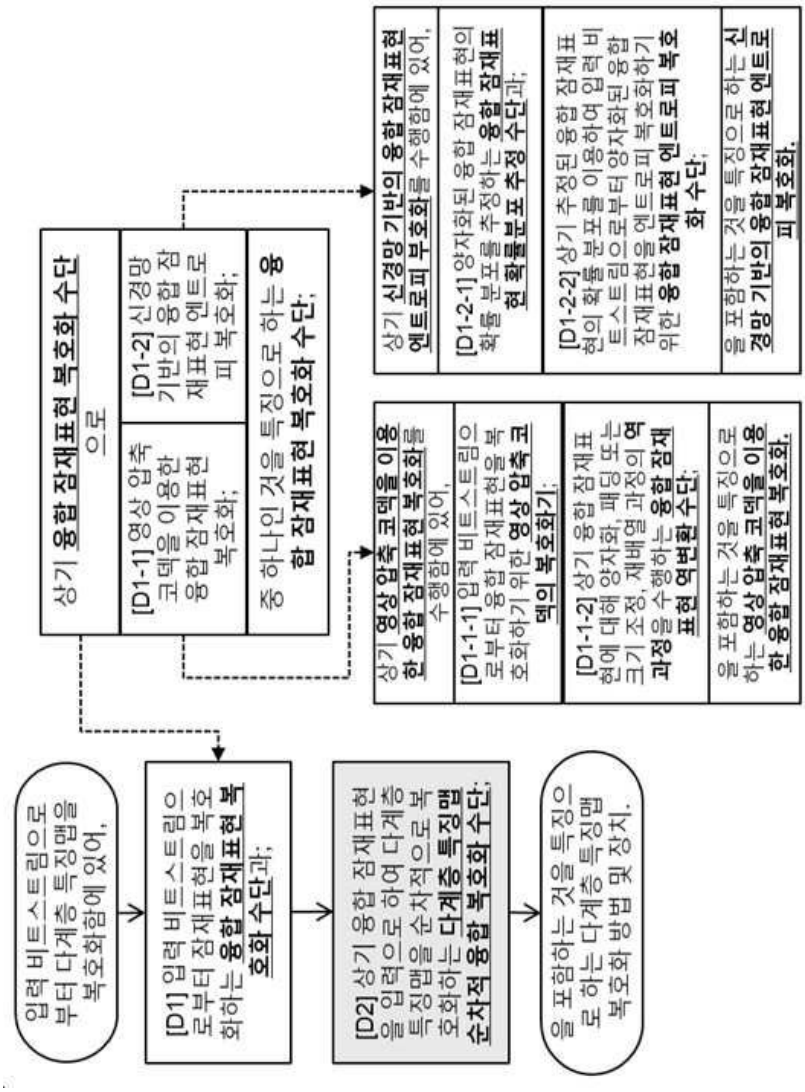
도면12


다계층 특징맵 순차적 융합 복호화 신경망
$y(=f_L)$
$\hat{F}_L$
$\hat{F}_{L-1}$
$\hat{F}_2$
$\hat{F}_1$
복호화 블록 2↑
특징맵 해상도의 순차적 증가
(계층 인덱스의 순차적 감소)

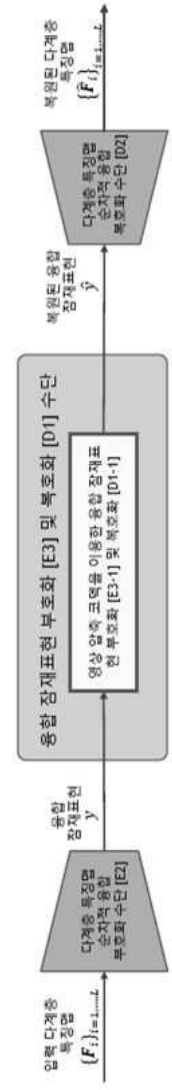
도면13


입력 영상으로부터 다계층 특징맵을 추출하여 부호화함에 있어,
[E1] 입력 영상으로부터 다계층 특징맵을 추출하는 **다계층 특징맵 추출 수단**과;
[E2] 상기 추출된 다계층 특징맵을 순차적으로 **융합 잠재표현으로 변환하여** 최종적으로 단일 계층의 **융합 잠재표현**을 출력하는 **다계층 특징맵 순차적 융합 부호화 수단**과;
[E2-1] 상기 추출된 다계층 특징맵은 Min-Max 정규화, Z-score 정규화 등의 통상적인 방법으로 정규화 되었을 수 있음
[E3] 상기 융합 잠재표현을 비트스트림으로 변환하는 **융합 잠재표현 부호화 수단**;
을 포함하는 것을 특징으로 하는 다계층 특징맵 부호화 방법 및 장치.
상기 **융합 잠재표현 부호화 수단**으로
[E3-1] 영상 압축 코덱을 이용한 융합 잠재표현 부호화;
[E3-2] 신경망 기반의 융합 잠재표현 엔트로피 부호화;
중 하나인 것을 특징으로 하는 **융합 잠재표현 부호화 수단**;
상기 **영상 압축 코덱을 이용한 융합 잠재표현 부호화**를 수행함에 있어,
[E3-1-1] 상기 융합 잠재표현에 대해 양자화, 패딩 또는 크기 조정, 재배열 등의 과정을 수행하여 부호화에 적합한 형식으로 만드는 **융합 잠재표현 변환 수단**;
[E3-1-2] 상기 변환된 융합 잠재표현을 비트스트림 형태로 부호화하기 위한 **영상 압축 코덱의 부호화기**;
을 포함하는 것을 특징으로 하는 **영상 압축 코덱을 이용한 융합 잠재표현 부호화**
상기 **신경망 기반의 융합 잠재표현 엔트로피 부호화**를 수행함에 있어,
[E3-2-1] 상기 융합 잠재표현을 양자화하는 **융합 잠재표현 양자화 수단**과;
[E3-2-2] 양자화된 융합 잠재표현의 확률분포를 추정하는 **융합 잠재표현 확률분포 추정 수단**과;
[E3-2-3] 상기 추정된 융합 잠재표현의 확률 분포를 이용하여 양자화된 융합 잠재표현을 엔트로피 부호화 하기 위한 **융합 잠재표현 엔트로피 부호화 수단**;
을 포함하는 것을 특징으로 하는 **신경망 기반의 융합 잠재표현 엔트로피 부호화**

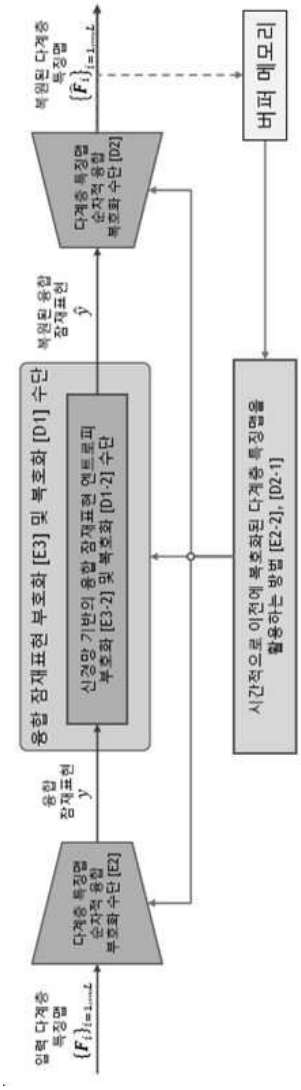
도면14


입력 비트스트림으로부터 다계층 특징맵을 복호화함에 있어,
[D1] 입력 비트스트림으로부터 잠재표현을 복호화하는 융합 잠재표현 복호화 수단과;
[D2] 상기 융합 잠재표현을 입력으로 하여 다계층 특징맵을 순차적으로 복호화하는 다계층 특징맵 순차적 융합 복호화 수단;
을 포함하는 것을 특징으로 하는 다계층 특징맵 복호화 방법 및 장치.
상기 융합 잠재표현 복호화 수단으로
[D1-1] 영상 압축 코덱을 이용한 융합 잠재표현 복호화;
[D1-2] 신경망 기반의 융합 잠재표현 엔트로피 복호화;
중 하나인 것을 특징으로 하는 융합 잠재표현 복호화 수단;
상기 영상 압축 코덱을 이용한 융합 잠재표현 복호화를 수행함에 있어,
[D1-1-1] 입력 비트스트림으로부터 융합 잠재표현을 복호화하기 위한 영상 압축 코덱의 복호화기;
[D1-1-2] 상기 융합 잠재표현에 대해 양자화, 패딩 또는 크기 조정, 재배열 과정의 역과정을 수행하는 융합 잠재표현 역변환 수단;
을 포함하는 것을 특징으로 하는 영상 압축 코덱을 이용한 융합 잠재표현 복호화.
상기 신경망 기반의 융합 잠재표현 엔트로피 복호화를 수행함에 있어,
[D1-2-1] 양자화된 융합 잠재표현의 확률 분포를 추정하는 융합 잠재표현 확률분포 추정 수단과;
[D1-2-2] 상기 추정된 융합 잠재표현의 확률 분포를 이용하여 입력 비트스트림으로부터 양자화된 융합 잠재표현을 엔트로피 복호화하기 위한 융합 잠재표현 엔트로피 복호화 수단;
을 포함하는 것을 특징으로 하는 신경망 기반의 융합 잠재표현 엔트로피 복호화.

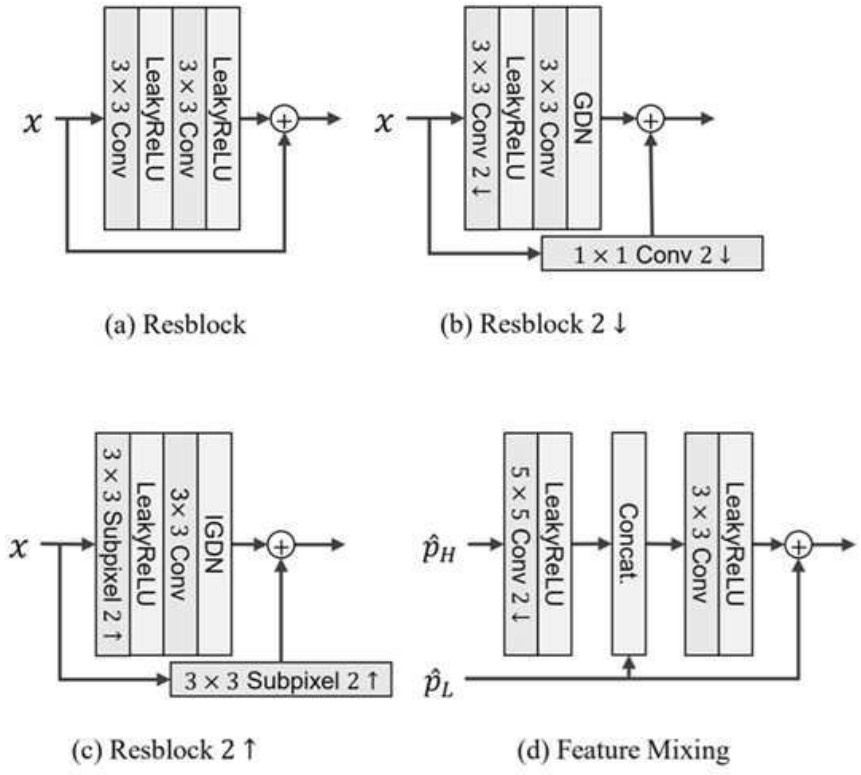
도면15


입력 다계층
특징맵
다계층 특징맵
순차적 융합
부호화 수단 [E2]
융합
잠재표현
y
융합 잠재표현 부호화 [E3] 및 복호화 [D1] 수단
영상 압축 코덱을 이용한 융합 잠재표
현 부호화 [E3-1] 및 복호화 [D1-1]
복원된 융합
잠재표현
ŷ
다계층특징맵
순차적 융합
복호화 수단 [D2]
복원된 다계층
특징맵

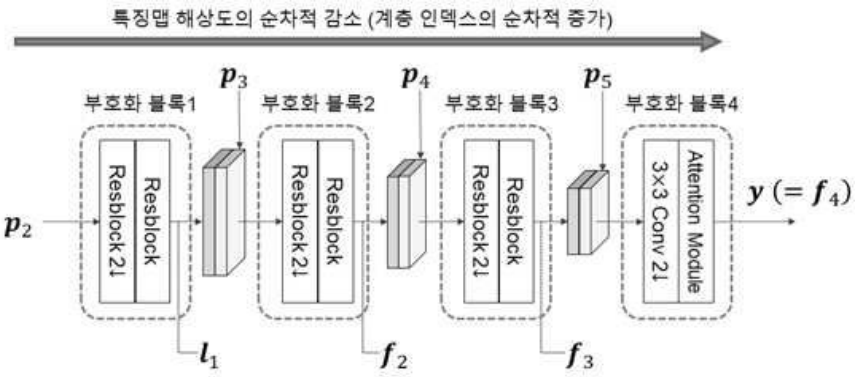
도면16


입력 다계층 특징맵
$\{F_i\}_{i=1,\ldots,L}$
다계층 특징맵 순차적 융합 부호화 수단 [E2]
융합 잠재표현 y
융합 잠재표현 부호화 [E3] 및 복호화 [D1] 수단
신경망 기반의 융합 잠재표현 엔트로피 부호화 [E3-2] 및 복호화 [D1-2] 수단
복원된 융합 잠재표현 $\hat{y}$
다계층 특징맵 순차적 융합 복호화 수단 [D2]
복원된 다계층 특징맵
$\{\hat{F}_i\}_{i=1,\ldots,L}$
버퍼 메모리
시간적으로 이전에 복호화된 다계층 특징맵을 활용하는 방법 [E2-2], [D2-1]

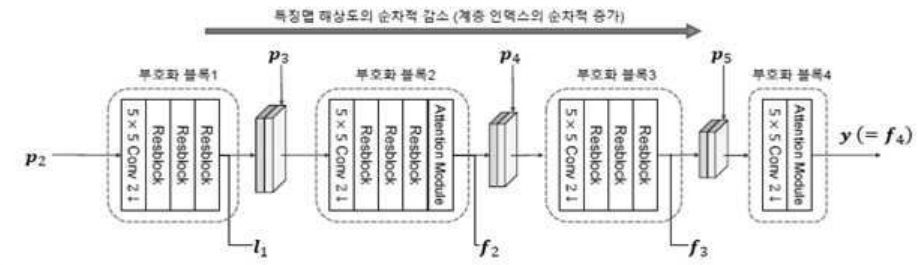
도면17



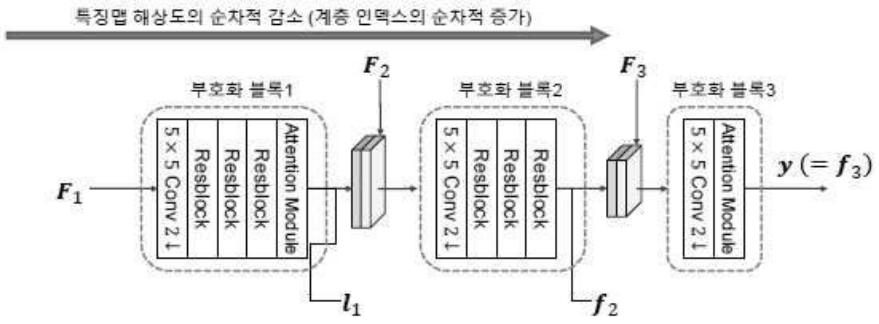
도면18



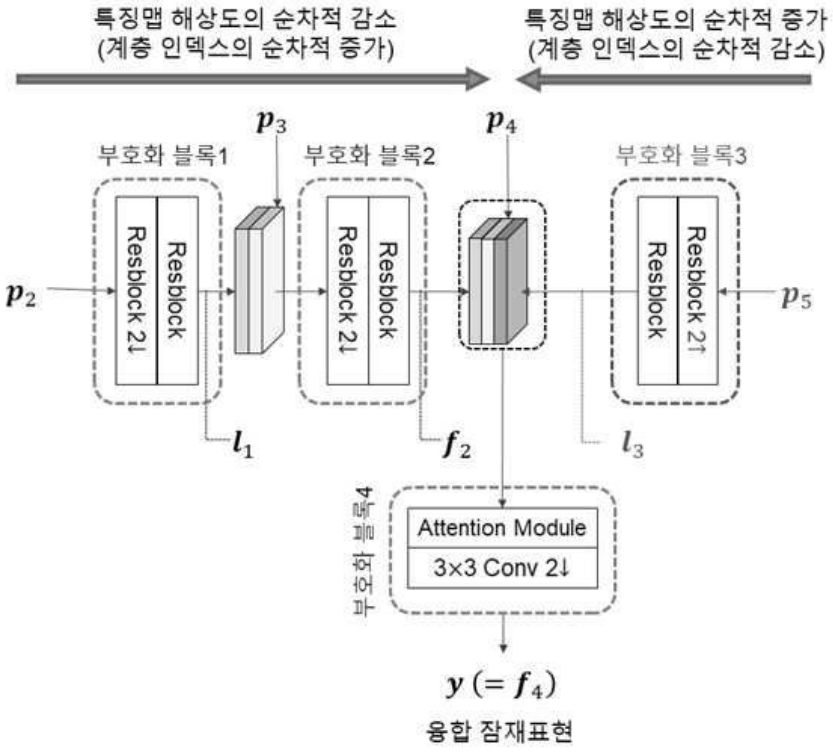
도면19



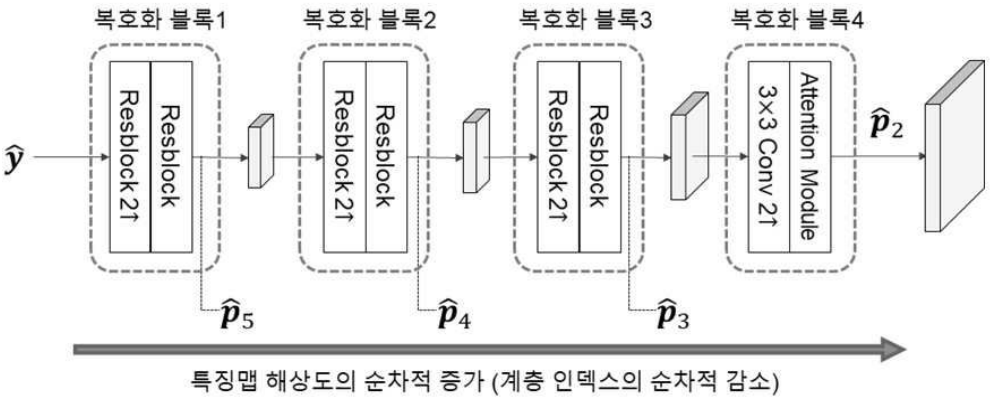
도면20



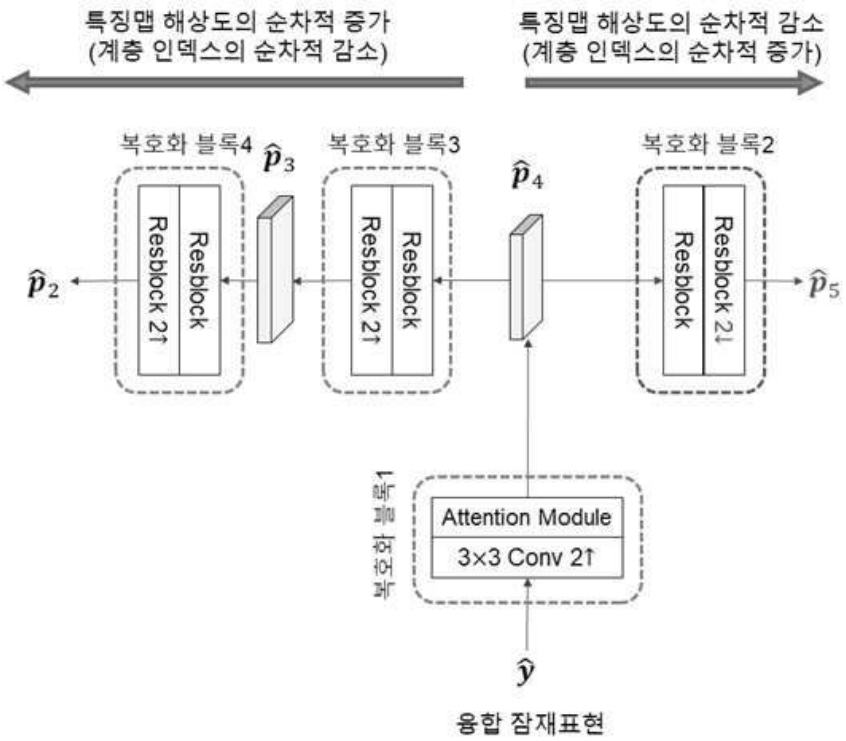
도면21



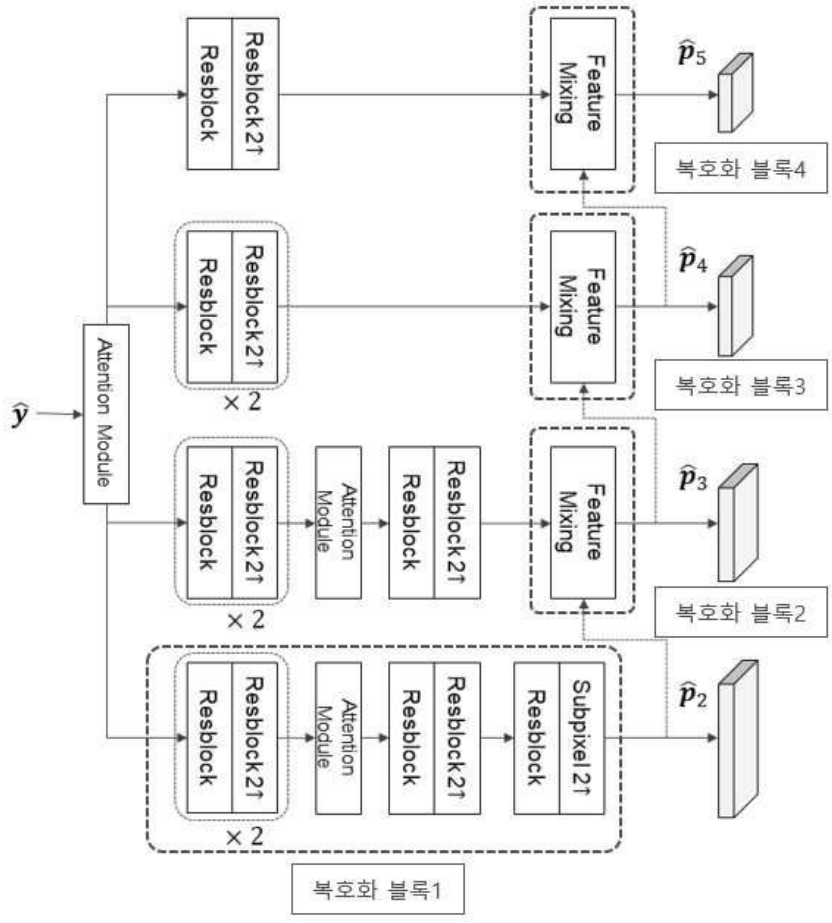
도면22



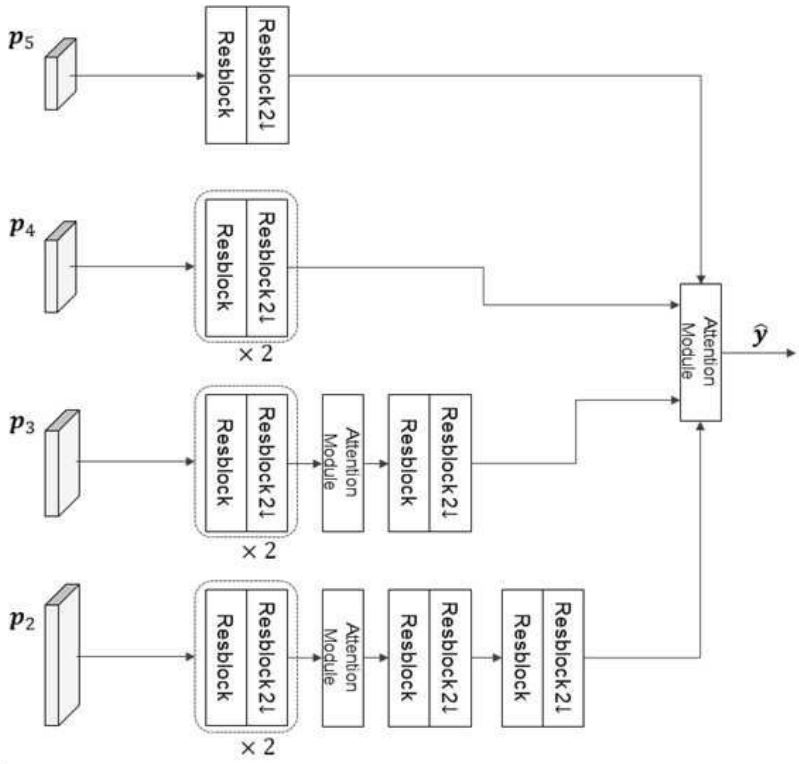
도면23


특징맵 해상도의 순차적 증가
(계층 인덱스의 순차적 감소)
특징맵 해상도의 순차적 감소
(계층 인덱스의 순차적 증가)
복호화 블록4
복호화 블록3
복호화 블록2
복호화 블록1
Resblock
Resblock 2↑
Resblock 2↓
Attention Module
3×3 Conv 2↑
융합 잠재표현

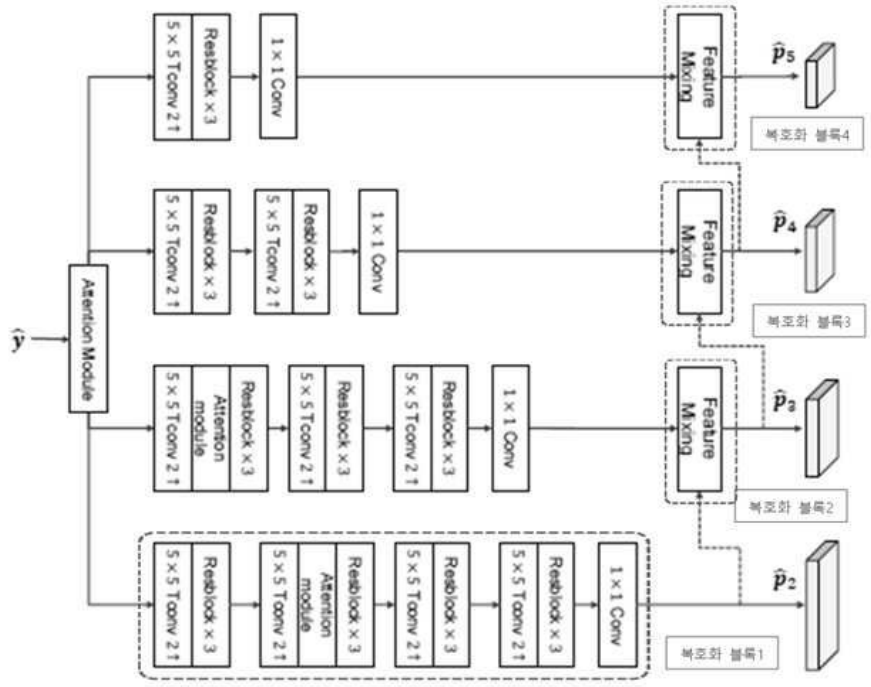
도면24


Resblock
Resblock 2↑
Feature Mixing
Attention Module
× 2
복호화 블록4
복호화 블록3
복호화 블록2
복호화 블록1
Subpixel 2↑

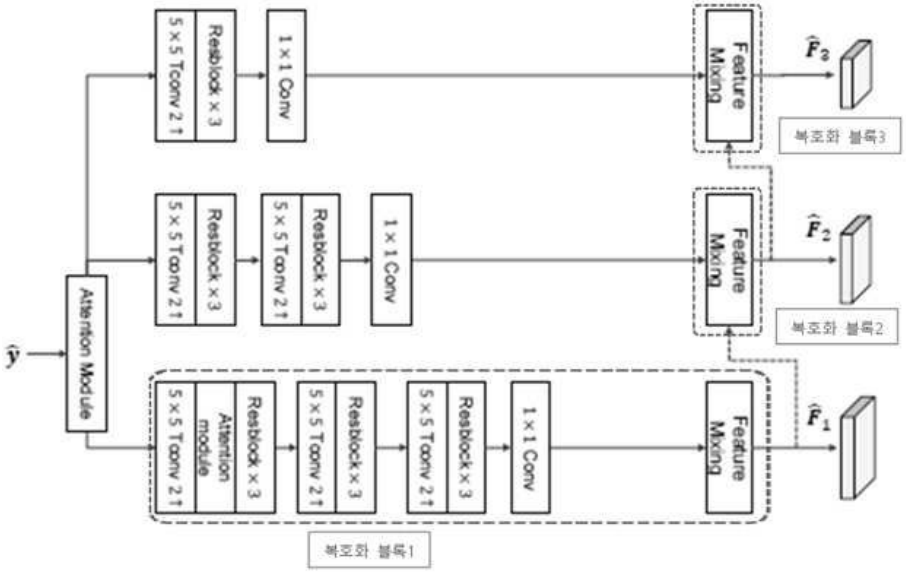
도면25


p_5
Resblock
Resblock2↓
p_4
Resblock
Resblock2↓
× 2
p_3
Resblock
Resblock2↓
× 2
Attention Module
Resblock
Resblock2↓
p_2
Resblock
Resblock2↓
× 2
Attention Module
Resblock
Resblock2↓
Resblock
Resblock2↓
Attention Module
$\hat{y}$

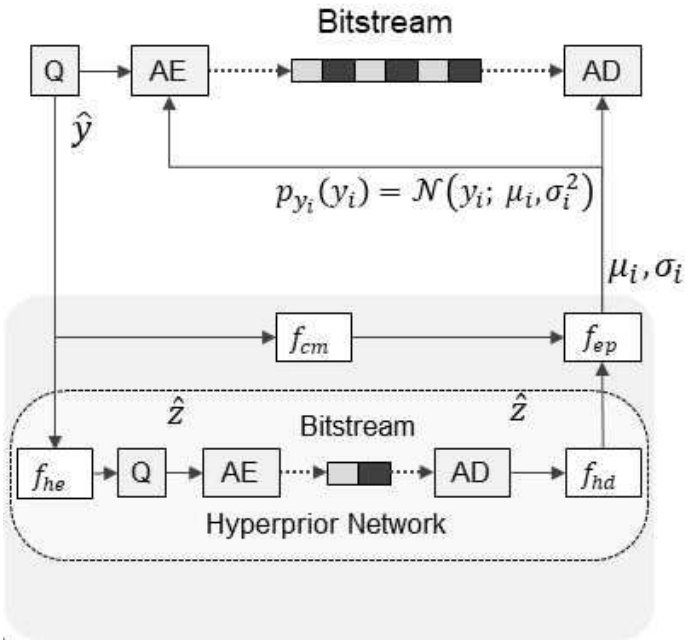
도면26


$\hat{y}$
Attention Module
5×5 Tconv 2↑
Resblock ×3
1×1 Conv
Feature Mixing
$\hat{p}_5$
복호화 블록4
$\hat{p}_4$
복호화 블록3
$\hat{p}_3$
복호화 블록2
$\hat{p}_2$
복호화 블록1
Attention module

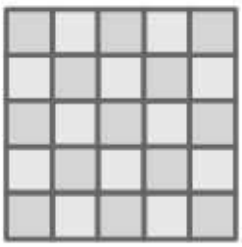
도면27


5×5 Tconv 2↑
Resblock ×3
1×1 Conv
Feature Mixing
복호화 블록3
Attention Module
Attention module
복호화 블록2
복호화 블록1

도면28


Bitstream
Q
AE
AD
$\hat{y}$
$p_{y_i}(y_i) = \mathcal{N}(y_i; \mu_i, \sigma_i^2)$
μ_i, σ_i
f_{cm}
f_{ep}
$\hat{z}$
Bitstream
f_{he}
Q
AE
AD
f_{hd}
Hyperprior Network

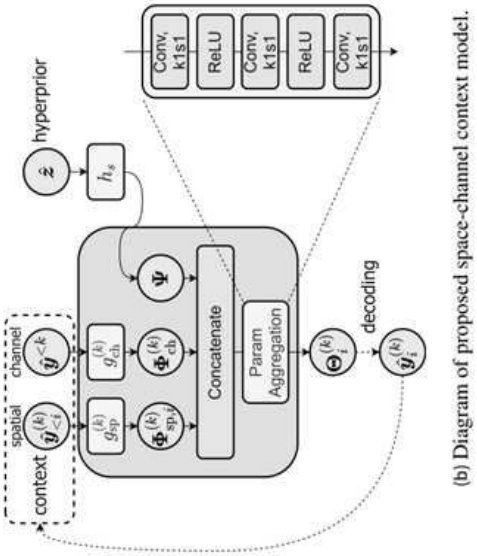
도면29



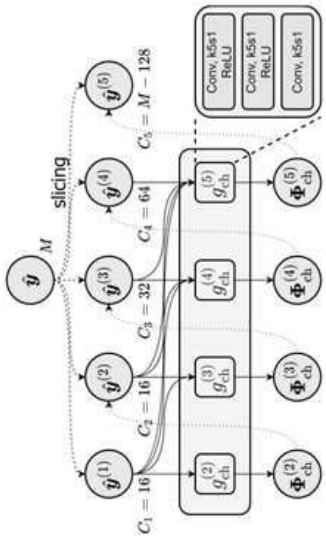
: checkerboard 패턴으로 나뉜 $\hat{y}^{(k)}$의
첫 번째 하위 그룹 $\hat{y}_1^{(k)}$의 공간적 위치 예시

: checkerboard 패턴으로 나뉜 $\hat{y}^{(k)}$의
두 번째 하위 그룹 $\hat{y}_2^{(k)}$의 공간적 위치 예시

도면30

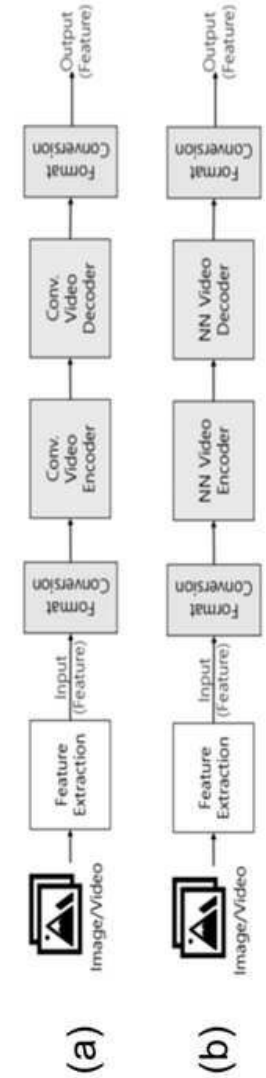


(b) Diagram of proposed space-channel context model.

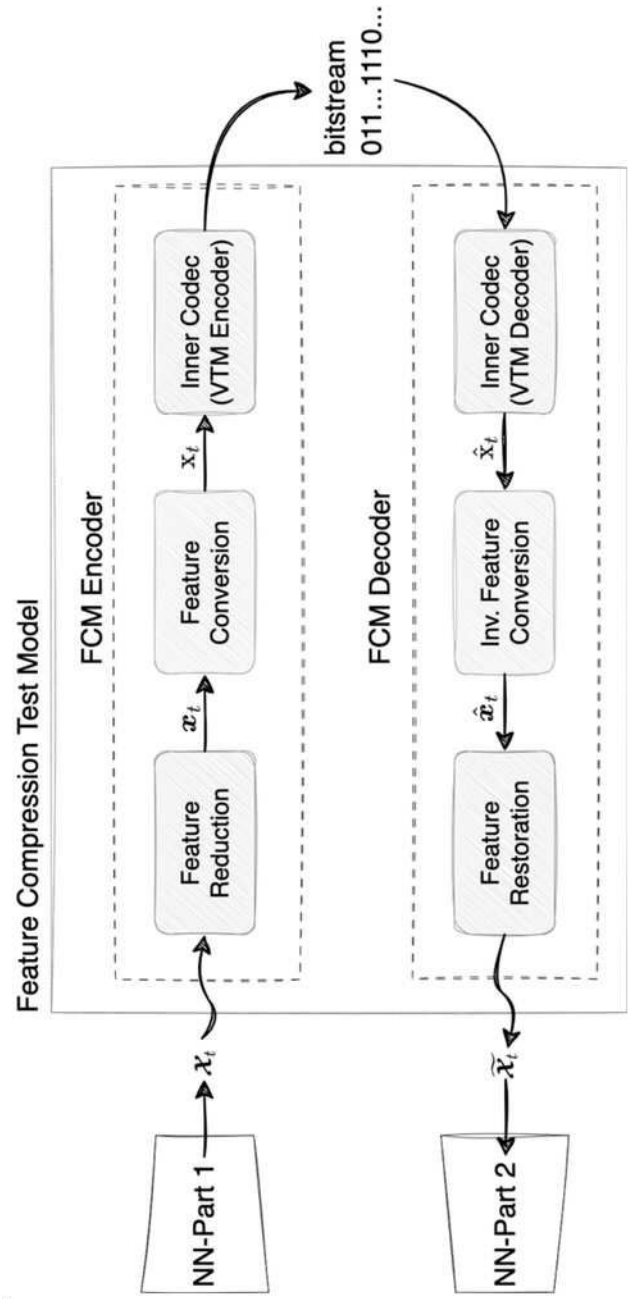


(a) Proposed uneven grouping for channel-conditional (CC) adaptive coding. The M-channel coding-symbols $\hat{\boldsymbol{y}}$ are grouped into 5 chunks with gradually increased number of channels C_k.

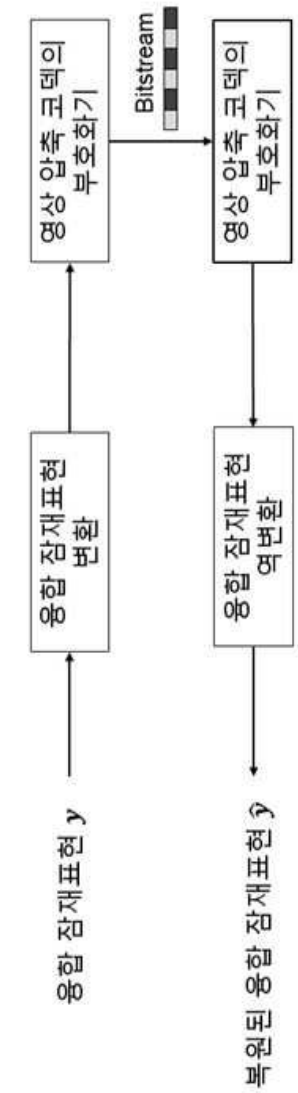
도면31


(a)
Image/Video
Feature Extraction
Input (Feature)
Format Conversion
Conv. Video Encoder
Conv. Video Decoder
Format Conversion
Output (Feature)
(b)
Image/Video
Feature Extraction
Input (Feature)
Format Conversion
NN Video Encoder
NN Video Decoder
Format Conversion
Output (Feature)

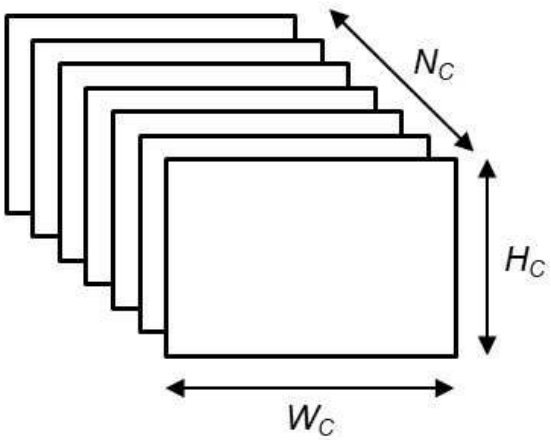
도면32


Feature Compression Test Model
FCM Encoder
Feature Reduction
Feature Conversion
Inner Codec (VTM Encoder)
bitstream
011...1110...
FCM Decoder
Feature Restoration
Inv. Feature Conversion
Inner Codec (VTM Decoder)
NN-Part 1
NN-Part 2

도면33

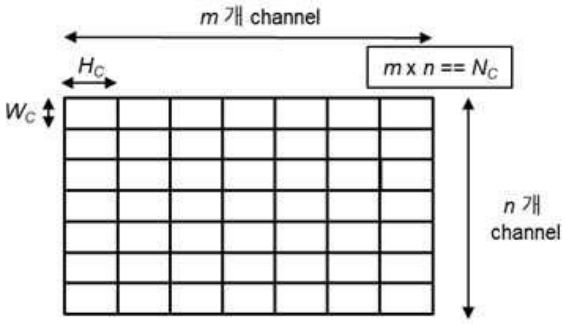


도면34

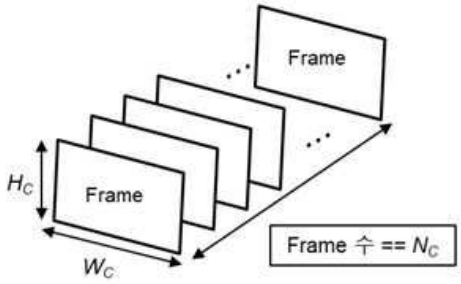


특징맵의 구성

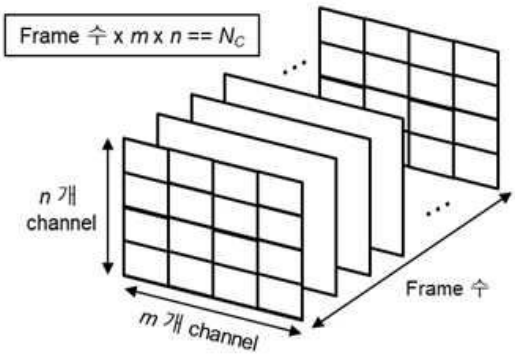
도면35



(a) 공간적(Spatial)으로 배열된 특징맵

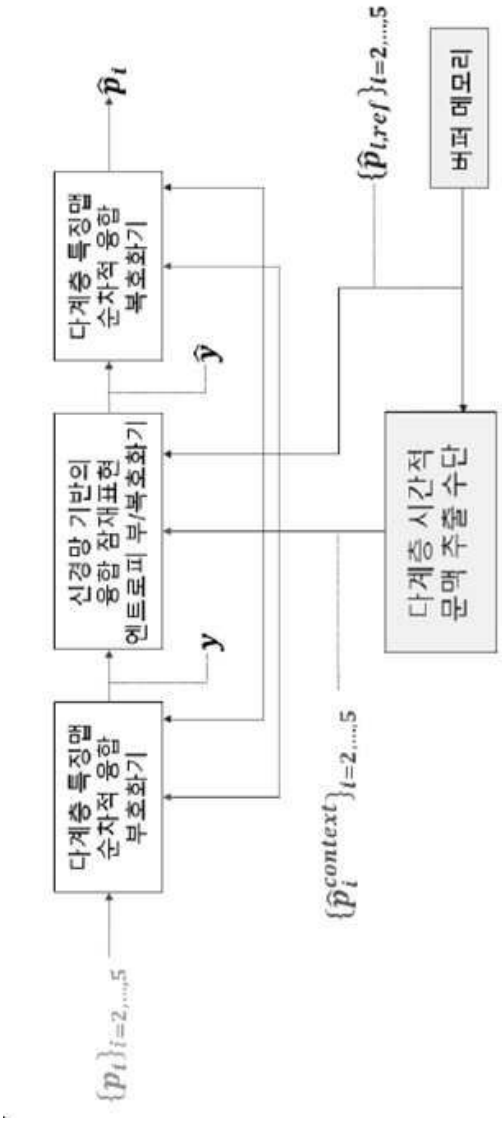


(b) 시간적(Temporal)으로 배열된 특징맵

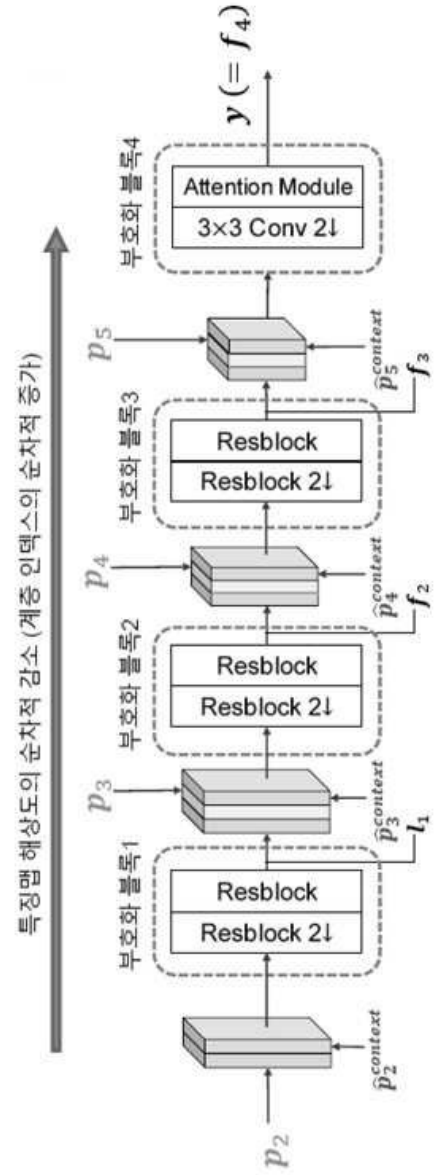


(c) 시공간적으로 배열된 특징맵

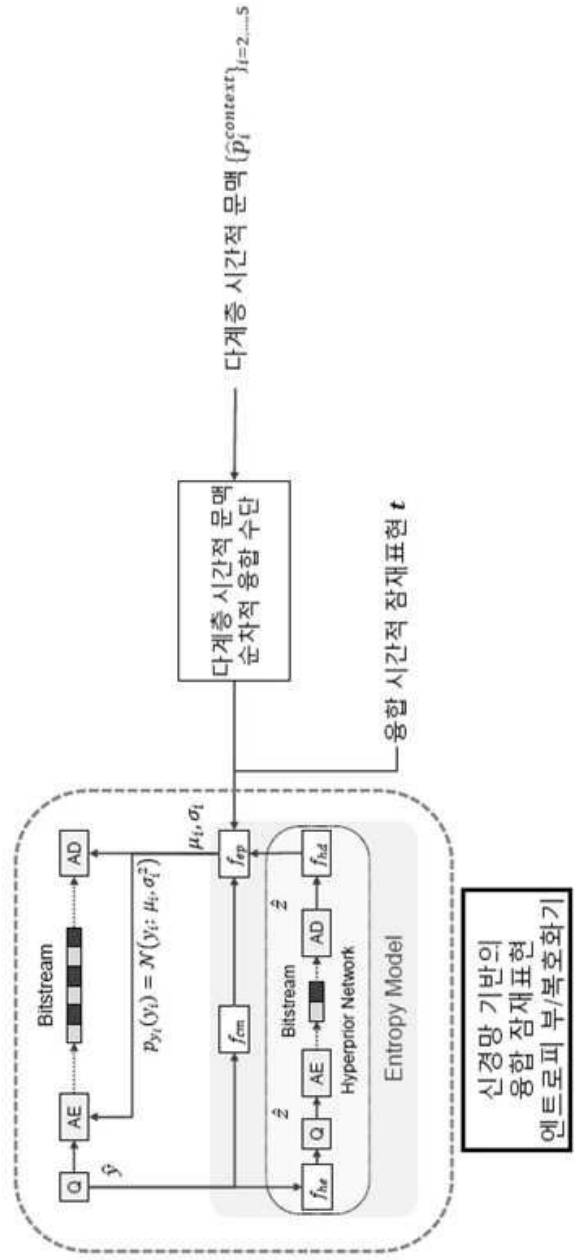
도면36


$\{p_i\}_{i=2,\ldots,5}$
다계층 특징맵 순차적 융합 부호화기
신경망 기반의 융합 잠재표현 엔트로피 부/복호화기
다계층 특징맵 순차적 융합 복호화기
$\hat{p}_i$
$\hat{y}$
y
$\{\hat{p}_i^{context}\}_{i=2,\ldots,5}$
다계층 시간적 문맥 추출 수단
버퍼 메모리
$\{\hat{p}_{i,ref}\}_{i=2,\ldots,5}$

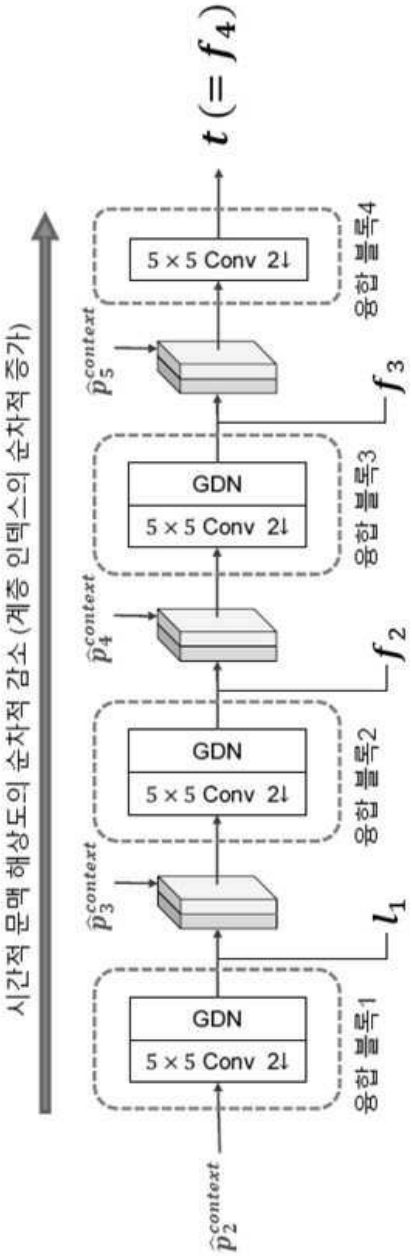
도면37


특징맵 해상도의 순차적 감소 (계층 인덱스의 순차적 증가)
$y\,(=f_4)$
부호화 블록4
Attention Module
3×3 Conv 2↓
부호화 블록3
Resblock
Resblock 2↓
부호화 블록2
Resblock
Resblock 2↓
부호화 블록1
Resblock
Resblock 2↓
p_5
p_4
p_3
p_2
$\hat{p}_5^{context}$
$\hat{p}_4^{context}$
$\hat{p}_3^{context}$
$\hat{p}_2^{context}$
f_3
f_2
l_1

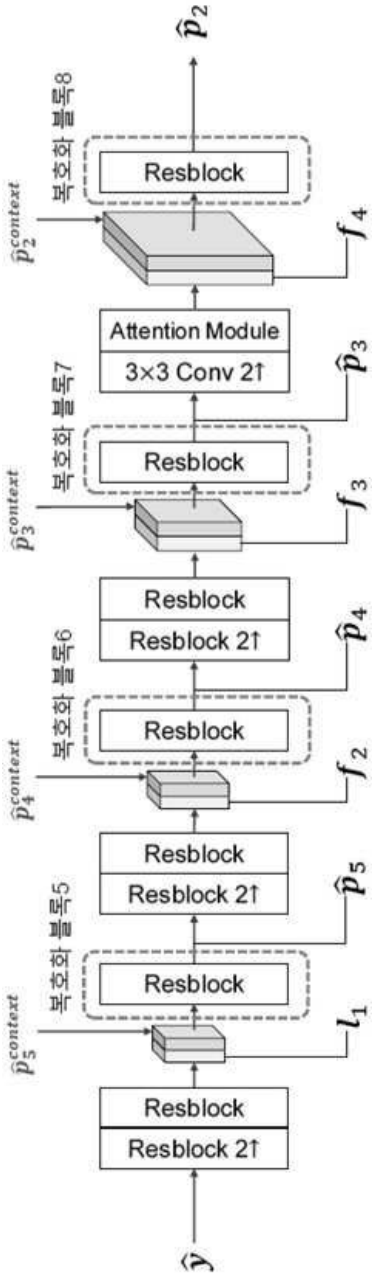
도면38


다계층 시간적 문맥 $\{\tilde{p}_i^{context}\}_{i=2,\ldots,S}$
다계층 시간적 문맥 순차적 융합 수단
융합 시간적 잠재표현 t
μ_i, σ_i
f_{ep}
f_{cm}
f_{hd}
f_{he}
AD
AE
Q
$\hat{y}$
$\hat{z}$
Bitstream
Hyperprior Network
Entropy Model
$p_{\hat{y}_i}(\hat{y}_i) = \mathcal{N}(\hat{y}_i; \mu_i, \sigma_i^2)$
신경망 기반의
융합 잠재표현
엔트로피 부/복호화기

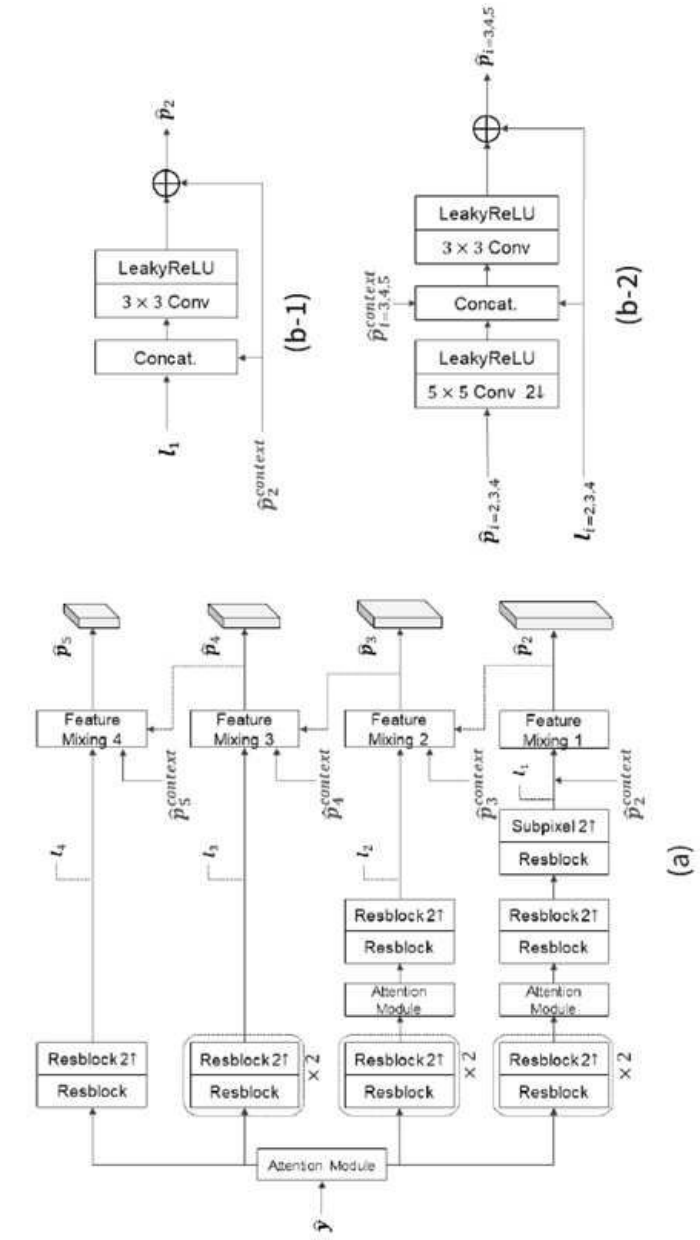
도면39


시간적 문맥 해상도의 순차적 감소 (계층 인덱스의 순차적 증가)
$\hat{p}_2^{context}$
융합 블록1
GDN
5 × 5 Conv 2↓
l_1
$\hat{p}_3^{context}$
융합 블록2
GDN
5 × 5 Conv 2↓
f_2
$\hat{p}_4^{context}$
융합 블록3
GDN
5 × 5 Conv 2↓
f_3
$\hat{p}_5^{context}$
융합 블록4
5 × 5 Conv 2↓
$t\ (= f_4)$

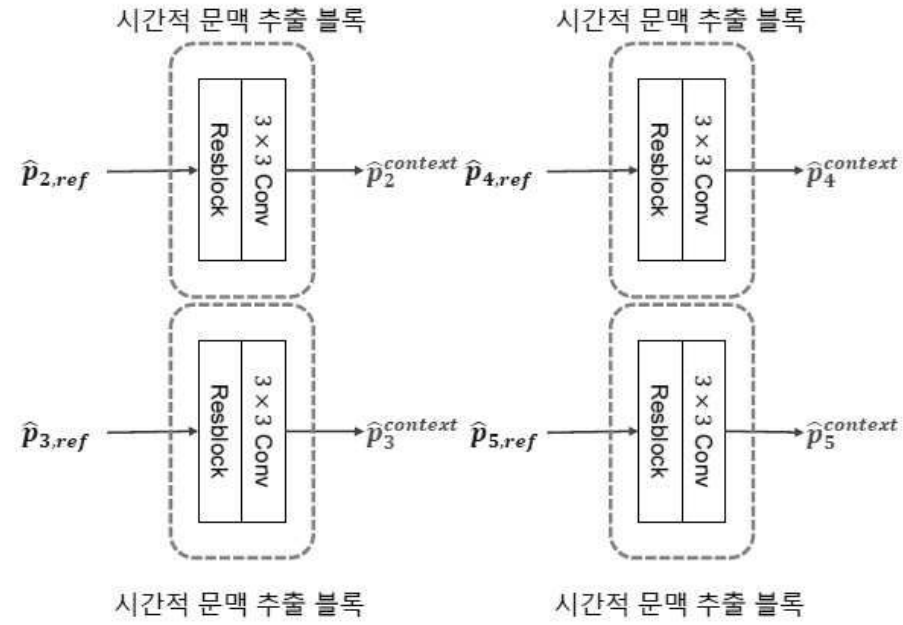
도면40


$\hat{p}_2$
복호화 블록8
Resblock
$\hat{p}_2^{context}$
f_4
Attention Module
3×3 Conv 2↑
$\hat{p}_3$
복호화 블록7
Resblock
$\hat{p}_3^{context}$
f_3
Resblock
Resblock 2↑
$\hat{p}_4$
복호화 블록6
Resblock
$\hat{p}_4^{context}$
f_2
Resblock
Resblock 2↑
$\hat{p}_5$
복호화 블록5
Resblock
$\hat{p}_5^{context}$
l_1
Resblock
Resblock 2↑
$\hat{y}$

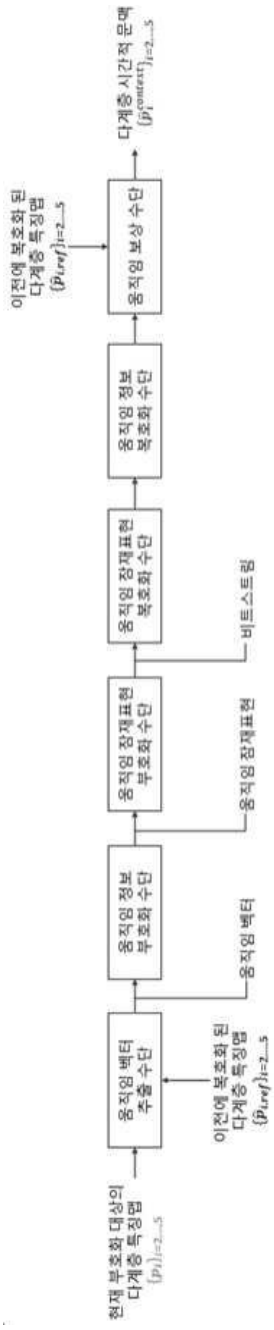
도면41


LeakyReLU
3 × 3 Conv
Concat.
(b-1)
LeakyReLU
3 × 3 Conv
Concat.
LeakyReLU
5 × 5 Conv 2↓
(b-2)
Feature Mixing 4
Feature Mixing 3
Feature Mixing 2
Feature Mixing 1
Subpixel 2↑
Resblock
Resblock 2↑
Resblock
Attention Module
× 2
Attention Module
(a)

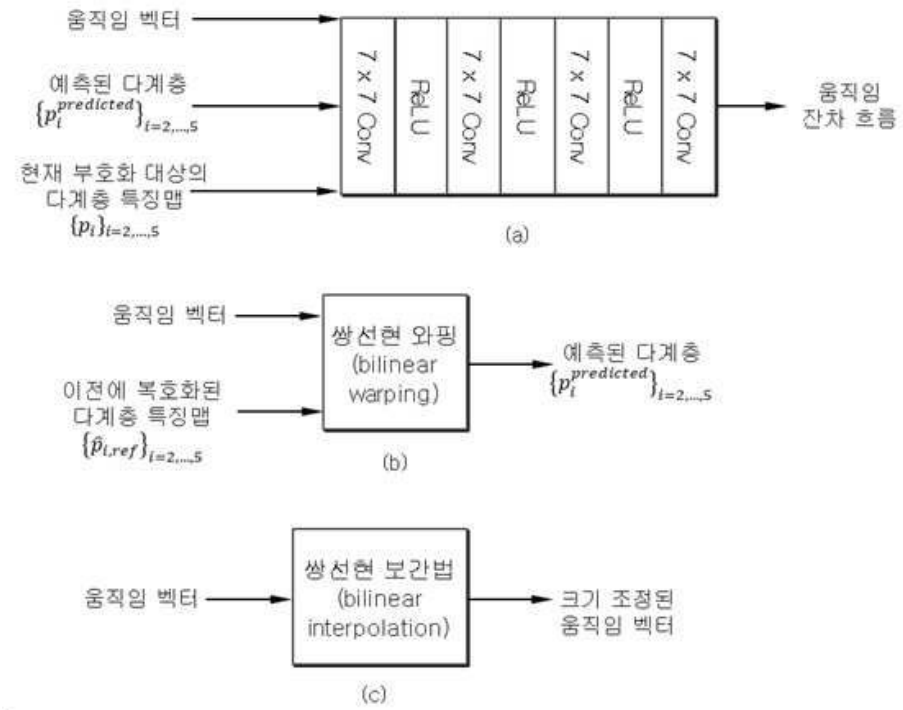
도면42


시간적 문맥 추출 블록
시간적 문맥 추출 블록
$\hat{p}_{2,ref}$
Resblock
3 × 3 Conv
$\hat{p}_2^{context}$
$\hat{p}_{4,ref}$
Resblock
3 × 3 Conv
$\hat{p}_4^{context}$
$\hat{p}_{3,ref}$
Resblock
3 × 3 Conv
$\hat{p}_3^{context}$
$\hat{p}_{5,ref}$
Resblock
3 × 3 Conv
$\hat{p}_5^{context}$
시간적 문맥 추출 블록
시간적 문맥 추출 블록

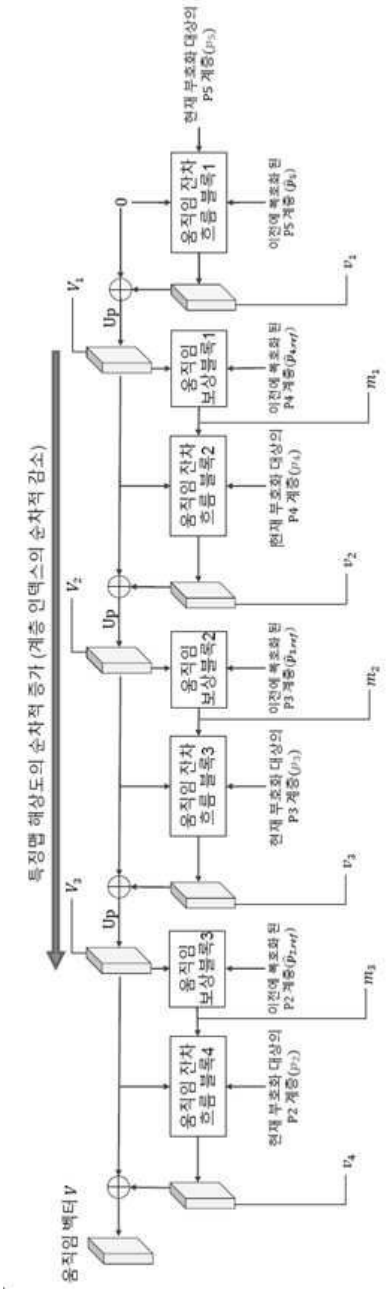
도면43


현재 부호화 대상의 다계층 특징맵
$\{p_i\}_{i=2...5}$
움직임 벡터 추출 수단
이전에 복호화 된 다계층 특징맵
$\{\hat{p}_{i,ref}\}_{i=2...5}$
움직임 정보 부호화 수단
움직임 벡터
움직임 잠재표현 부호화 수단
움직임 잠재표현
움직임 잠재표현 복호화 수단
비트스트림
움직임 정보 복호화 수단
움직임 보상 수단
이전에 복호화 된 다계층 특징맵
$\{\hat{p}_{i,ref}\}_{i=2...5}$
다계층 시간적 문맥
$\{\hat{p}_i^{context}\}_{i=2...5}$

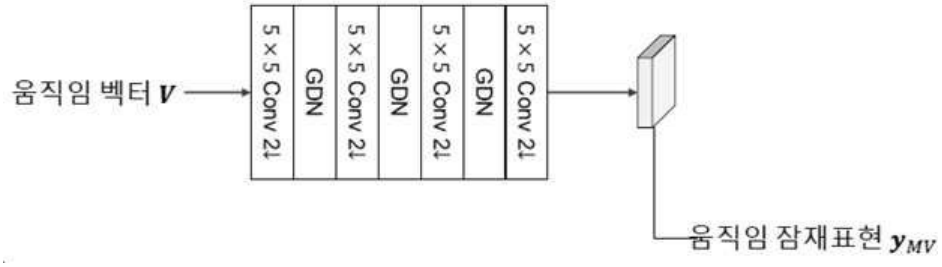
도면44


움직임 벡터
예측된 다계층
$\{p_i^{predicted}\}_{i=2,...,5}$
현재 부호화 대상의
다계층 특징맵
$\{p_i\}_{i=2,...,5}$
7 x 7 Conv
ReLU
7 x 7 Conv
ReLU
7 x 7 Conv
ReLU
7 x 7 Conv
움직임
잔차 흐름
(a)
움직임 벡터
이전에 복호화된
다계층 특징맵
$\{\hat{p}_{i,ref}\}_{i=2,...,5}$
쌍선현 와핑
(bilinear
warping)
예측된 다계층
$\{p_i^{predicted}\}_{i=2,...,5}$
(b)
움직임 벡터
쌍선현 보간법
(bilinear
interpolation)
크기 조정된
움직임 벡터
(c)

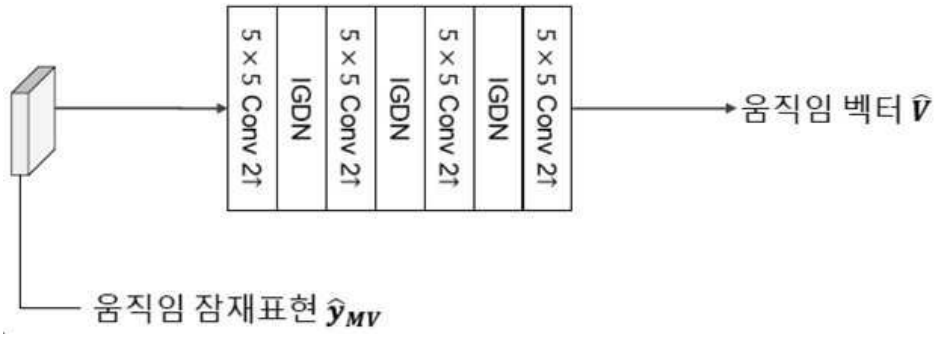
도면45


특징맵 해상도의 순차적 증가 (계층 인덱스의 순차적 감소)
움직임 벡터 V
움직임 잔차 흐름 블록1
움직임 잔차 흐름 블록2
움직임 잔차 흐름 블록3
움직임 잔차 흐름 블록4
움직임 보상블록1
움직임 보상블록2
움직임 보상블록3
Up
0
v_1
v_2
v_3
v_4
m_1
m_2
m_3

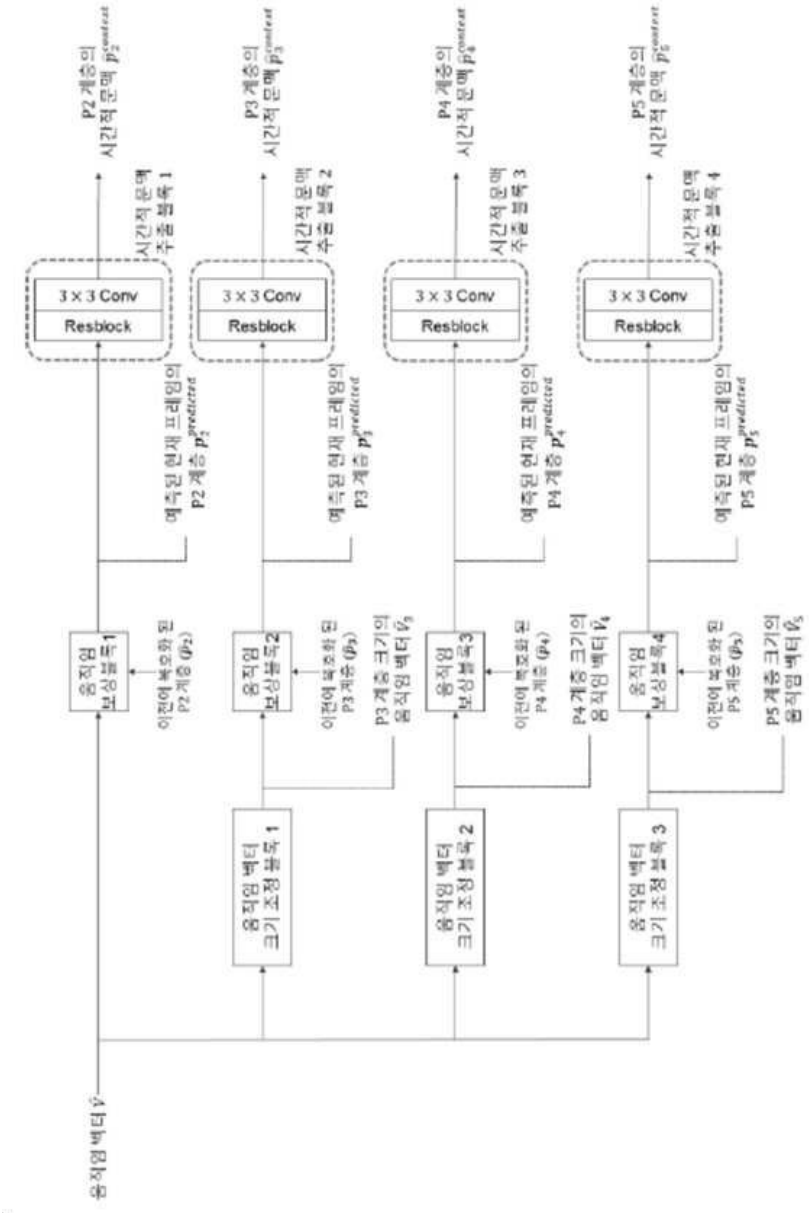
도면46


움직임 벡터 V
5 × 5 Conv 21
GDN
5 × 5 Conv 21
GDN
5 × 5 Conv 21
GDN
5 × 5 Conv 21
움직임 잠재표현 y_{MV}

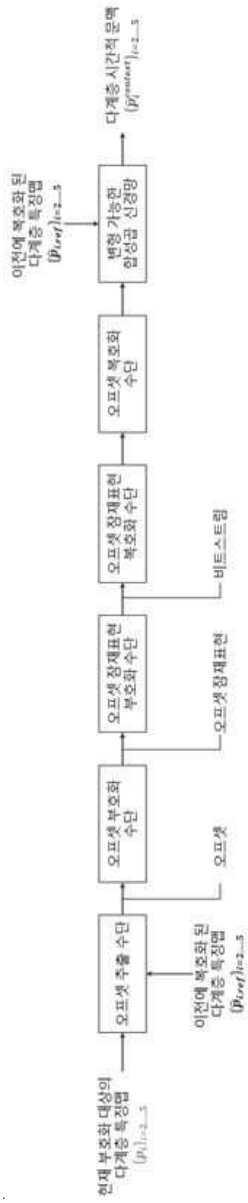
도면47


5 × 5 Conv 2↓
IGDN
5 × 5 Conv 2↓
IGDN
5 × 5 Conv 2↓
IGDN
5 × 5 Conv 2↓
움직임 벡터 $\hat{v}$
움직임 잠재표현 $\hat{y}_{MV}$

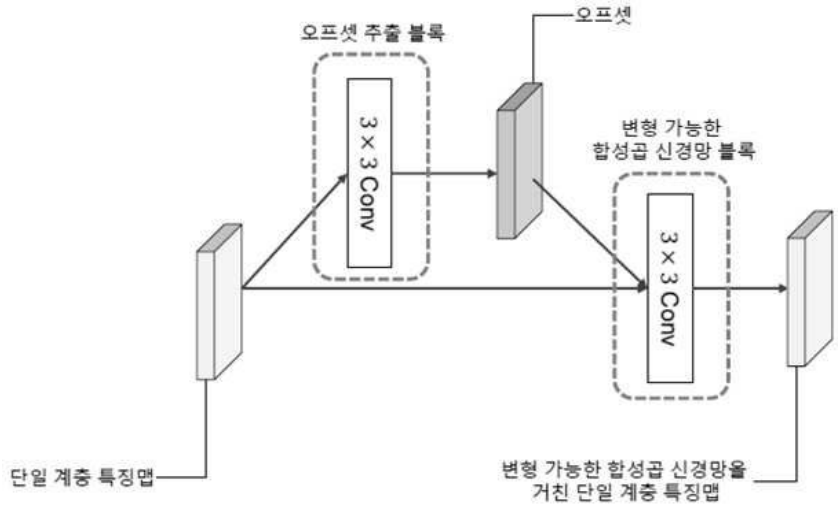
도면48


3 × 3 Conv
Resblock
움직임 보상블록1
움직임 보상블록2
움직임 보상블록3
움직임 보상블록4
움직임 벡터 크기 조정 블록 1
움직임 벡터 크기 조정 블록 2
움직임 벡터 크기 조정 블록 3

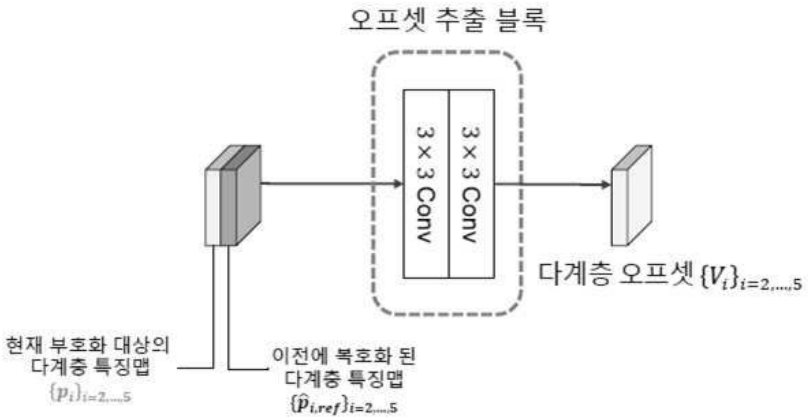
도면49


현재 부호화 대상의 다계층 특징맵
오프셋 추출 수단
이전에 복호화 된 다계층 특징맵
오프셋 부호화 수단
오프셋
오프셋 잠재표현 부호화 수단
오프셋 잠재표현
오프셋 잠재표현 복호화 수단
비트스트림
오프셋 복호화 수단
변형 가능한 합성곱 신경망
이전에 복호화 된 다계층 특징맵
다계층 시간적 문맥

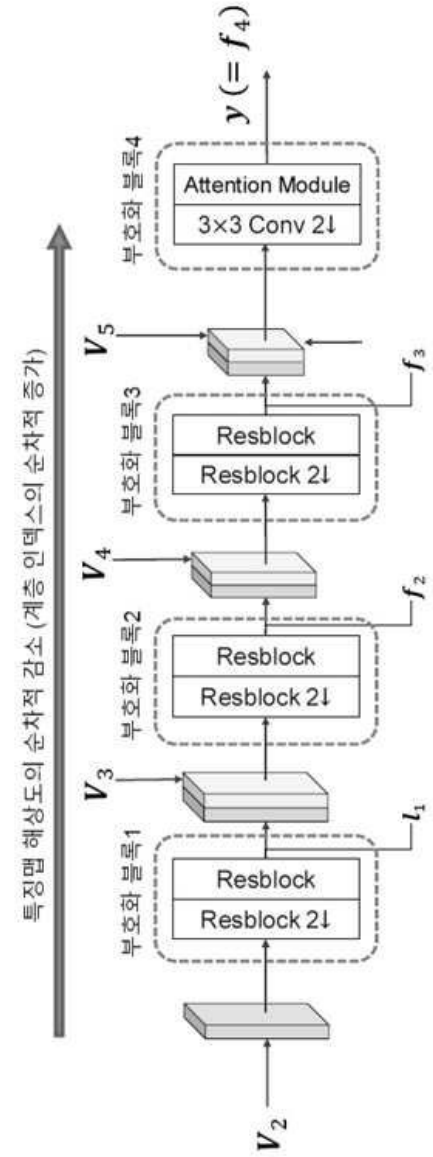
도면50


오프셋 추출 블록
오프셋
3 × 3 Conv
변형 가능한
합성곱 신경망 블록
3 × 3 Conv
단일 계층 특징맵
변형 가능한 합성곱 신경망을
거친 단일 계층 특징맵

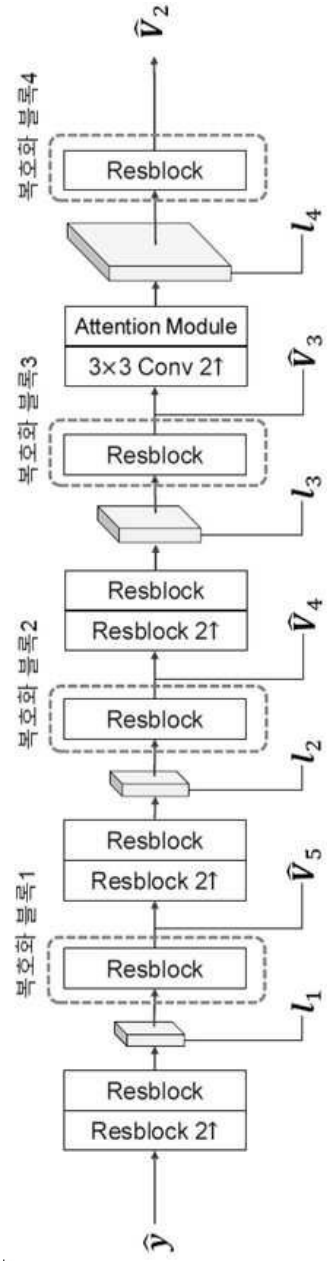
도면51


오프셋 추출 블록
3 × 3 Conv
3 × 3 Conv
다계층 오프셋 $\{V_i\}_{i=2,\ldots,5}$
현재 부호화 대상의
다계층 특징맵
$\{P_i\}_{i=2,\ldots,5}$
이전에 복호화 된
다계층 특징맵
$\{\hat{P}_{i,ref}\}_{i=2,\ldots,5}$

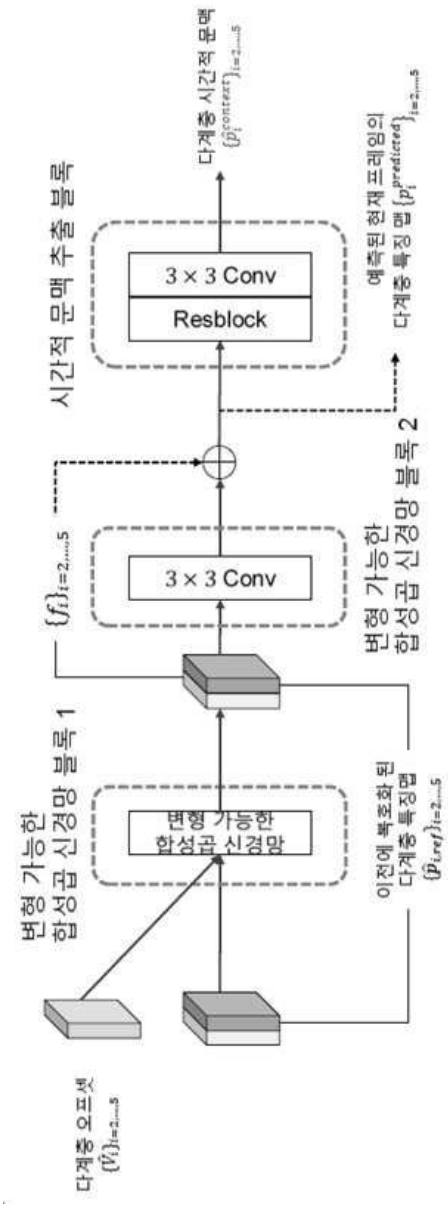
도면52


$y\ (=f_4)$
부호화 블록4
Attention Module
3×3 Conv 2↓
V_5
f_3
부호화 블록3
Resblock
Resblock 2↓
V_4
f_2
부호화 블록2
Resblock
Resblock 2↓
V_3
l_1
부호화 블록1
Resblock
Resblock 2↓
V_2
특징맵 해상도의 순차적 감소 (계층 인덱스의 순차적 증가)

도면53


$\hat{v}_2$
복호화 블록4
Resblock
l_4
Attention Module
3×3 Conv 2↑
$\hat{v}_3$
복호화 블록3
Resblock
l_3
Resblock
Resblock 2↑
$\hat{v}_4$
복호화 블록2
Resblock
l_2
Resblock
Resblock 2↑
$\hat{v}_5$
복호화 블록1
Resblock
l_1
Resblock
Resblock 2↑
$\hat{y}$

도면54


다계층 시간적 문맥
시간적 문맥 추출 블록
3 × 3 Conv
Resblock
예측된 현재 프레임의
다계층 특징 맵
변형 가능한
합성곱 신경망 블록 2
3 × 3 Conv
변형 가능한
합성곱 신경망 블록 1
변형 가능한
합성곱 신경망
이전에 복호화 된
다계층 특징맵
다계층 오프셋