

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구  
국제사무국



(10) 국제공개번호  
**WO 2020/242260 A1**

2020년 12월 3일 (03.12.2020)

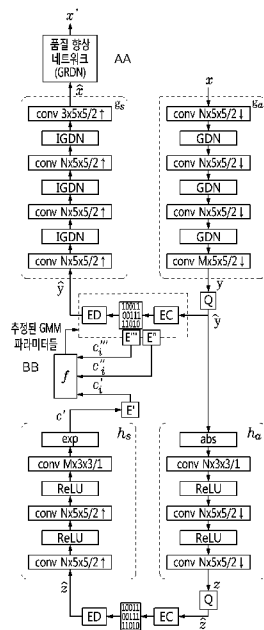
- (51) 국제특허분류: *H04N 1/41* (2006.01) *G06N 20/00* (2019.01)
- (21) 국제출원번호: PCT/KR2020/007039
- (22) 국제출원일: 2020년 5월 29일 (29.05.2020)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2019-0064882 2019년 5월 31일 (31.05.2019) KR  
10-2020-0065289 2020년 5월 29일 (29.05.2020) KR
- (71) 출원인: 한국전자통신연구원 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) [KR/KR]; 34129 대전광역시 유성구 가정로 218 (가정동), Daejeon (KR).
- (72) 발명자: 이주영 (LEE, Joo-Young); 34166 대전광역시 유성구 문화원로 13, 107동 804호, Daejeon (KR). 조승현 (CHO, Seung-Hyun); 34020 대전광역시 유성구 배울2로 61, 1014동 1204호, Daejeon (KR). 고현석 (KO, Hyun-suk); 34061 대전광역시 유성구 송림로 13, 101동 301호, Daejeon (KR). 권형진 (KWON, Hyoung-Jin); 30126 세종특별자치시 새롬북로 14, 303동 1304호, Sejong-si (KR). 김연희 (KIM, Youn-Hee); 34130 대전광역시 유성

구 가정로 270, 1동 323호, Daejeon (KR). 김종호 (KIM, Jong-Ho); 34070 대전광역시 유성구 지족북로 60, 205동 2001호, Daejeon (KR). 정세윤 (JEONG, Se-Yoon); 34070 대전광역시 유성구 지족북로 60, 207동 1401호, Daejeon (KR). 김휘용 (KIM, Hui-Yong); 34090 대전광역시 유성구 은구비남로 34, 810동 201호, Daejeon (KR). 최진수 (CHOI, Jin-Soo); 34067 대전광역시 유성구 반석서로 98, 609동 1605호, Daejeon (KR).

- (74) 대리인: 한양특허법인 (HANYANG PATENT FIRM); 06296 서울시 강남구 논현로38길 12, 한양빌딩, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: METHOD AND DEVICE FOR MACHINE LEARNING-BASED IMAGE COMPRESSION USING GLOBAL CONTEXT

(54) 발명의 명칭: 전역적 문맥을 이용하는 기계 학습 기반의 이미지 압축을 위한 방법 및 장치



AA ... Quality enhancement network (GRDN)  
BB ... Estimated GMM parameters

(57) Abstract: Provided are a method and device for machine learning-based image compression using global context. A disclosed image compression network employs an existing image quality enhancement network for an end-to-end joint learning scheme. The image compression network can jointly optimize image compression and quality enhancement. Image compression networks and image enhancement networks can be easily combined within an integrated architecture that minimizes total loss, and can be easily jointed and optimized.

(57) 요약서: 전역적 문맥을 이용하는 기계 학습 기반의 이미지 압축을 위한 방법 및 장치가 제공된다. 개시되는 이미지 압축 네트워크는 엔드-투-엔드 조인트 학습 스킴에 대하여 기존의 이미지 품질 향상 네트워크를 채용한다. 이미지 압축 네트워크는 이미지 압축 및 품질 향상을 조인트로 최적화할 수 있다. 이미지 압축 네트워크들 및 이미지 향상 네트워크들은 총 손실을 최소화하는 통합된 아키텍처 내에서 용이하게 조합될 수 있고, 용이하게 조인트되어 최적화될 수 있다.

WO 2020/242260 A1

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

규칙 4.17에 의한 선언서:

— 신규성을 해치지 아니하는 개시 또는 신규성 상실의 예외에 관한 선언 (규칙 4.17(v))

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

## 명세서

### 발명의 명칭: 전역적 문맥을 이용하는 기계 학습 기반의 이미지 압축을 위한 방법 및 장치

#### 기술분야

- [1] 아래의 실시예들은 비디오의 복호화 방법, 복호화 장치, 부호화 방법 및 부호화 장치에 관한 것으로서, 전역적 문맥을 이용하는 기계 학습에 기반하여 이미지에 대한 압축을 제공하는 복호화 방법, 복호화 장치, 부호화 방법 및 부호화 장치에 관한 것이다.
- [2]본 발명은 2019년 5월 31일 출원된 한국특허출원 제10-2019-0064882호의 출원일의 이익을 주장하며, 그 내용 전부는 본 명세서에 포함된다.
- [3]본 발명은 2020년 5월 29일 출원된 한국특허출원 제10-2020-0065289호의 출원일의 이익을 주장하며, 그 내용 전부는 본 명세서에 포함된다.

#### 배경기술

- [4]최근에는 학습된 이미지 압축 방법들이 활발하게 연구되고 있다. 이러한 학습된 이미지 압축 방법들 중 엔트로피-최소화(entropy-minimization) 기반의 접근방법(approach)들은 BPG 및 JPEG2000과 같은 통상적인 이미지 코덱들에 비해 우월한 결과들을 달성하였다.
- [5]그러나, 이미지 압축의 처리에서 품질 향상(quality enhancement) 및 율-최적화(rate-minimization)는 상충하여 커플된다. 즉, 고 이미지 품질의 유지는 낮은 압축률을 수반하고, 그 반대도 마찬가지이다.
- [6]그러나, 이미지 압축과 함께하여 별도의 품질 향상을 공동으로 훈련함으로써, 코딩 효율이 향상될 수 있다.

#### 발명의 상세한 설명

##### 기술적 과제

- [7]일 실시예는 전역적 문맥을 이용하는 기계 학습에 기반하여 이미지에 대한 압축을 제공하는 부호화 장치, 부호화 방법, 복호화 장치 및 복호화 방법을 제공할 수 있다.

##### 과제 해결 수단

- [8]일 측에 있어서, 입력 이미지에 대해 엔트로피 모델을 사용하는 엔트로피 부호화를 수행하여 비트스트림을 생성하는 단계; 및 상기 비트스트림을 전송 또는 저장하는 단계를 포함하는 부호화 방법이 제공된다.
- [9]상기 엔트로피 모델은 문맥-적응형 엔트로피 모델일 수 있다.
- [10]상기 문맥-적응형 엔트로피 모델은 문맥들의 서로 상이한 3개의 타입들을 활용할 수 있다.
- [11]상기 문맥들은 가우시안 혼합 모델의 파라미터를 추정하기 위해 사용될 수 있다.

- [12] 상기 파라미터는 가중치 파라미터, 평균 파라미터 및 표준 편차 파라미터를 포함할 수 있다.
- [13] 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델일 수 있다,
- [14] 상기 문맥-적응형 엔트로피 모델은 전역 문맥을 사용할 수 있다.
- [15] 상기 엔트로피 부호화는 이미지 압축 네트워크 및 품질 향상 네트워크의 결합에 의해 수행될 수 있다.
- [16] 상기 품질 향상 네트워크는 매우 깊은 슈퍼 레졸루션(Very Deep Super Resolution; VDSR), 잔차 밀도 네트워크(Residual Dense Network; RDN) 또는 그룹된 잔차 밀도 네트워크(Grouped Residual Dense Network; GRDN)일 수 있다.
- [17] 상기 입력 이미지에 수평 방향의 패딩 또는 수직 방향의 패딩이 적용될 수 있다.
- [18] 상기 수평 방향의 패딩은 상기 입력 이미지의 수직 축 상의 중심에 하나 이상의 행들을 삽입하는 것일 수 있다.
- [19] 상기 수직 방향의 패딩은 상기 입력 이미지의 수평 축 상의 중심에 하나 이상의 열들을 삽입하는 것일 수 있다.
- [20] 상기 수평 방향의 패딩은 상기 입력 이미지의 높이가  $k$ 의 배수가 아닐 경우에 수행될 수 있다.
- [21] 상기 수직 방향의 패딩은 상기 입력 이미지의 폭이  $k$ 의 배수가 아닐 경우에 수행될 수 있다
- [22] 상기  $k$ 는  $2^n$ 이고,
- [23] 상기  $n$ 은 상기 입력 이미지에 대한 다운-스케일링들의 개수일 수 있다.
- [24] 상기 부호화 방법에 의하여 생성된 상기 비트스트림을 기록하는 기록 매체가 제공될 수 있다.
- [25] 다른 일 측에 있어서, 비트스트림을 획득하는 통신부; 및 상기 비트스트림에 대해 엔트로피 모델을 사용하는 복호화를 수행하여 재구성된 이미지를 생성하는 처리부를 포함하는 복호화 장치가 제공된다.
- [26] 또 다른 일 측에 있어서, 비트스트림을 획득하는 단계; 및 상기 비트스트림에 대해 엔트로피 모델을 사용하는 복호화를 수행하여 재구성된 이미지를 생성하는 단계를 포함하는 복호화 방법이 제공된다.
- [27] 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델일 수 있다.
- [28] 상기 문맥-적응형 엔트로피 모델은 문맥들의 서로 상이한 3 개의 타입들을 활용할 수 있다.
- [29] 상기 문맥들은 가우시안 혼합 모델의 파라미터를 추정하기 위해 사용될 수 있다.
- [30] 상기 파라미터는 가중치 파라미터, 평균 파라미터 및 표준 편차 파라미터를 포함할 수 있다.
- [31] 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델일 수 있다.
- [32] 상기 문맥-적응형 엔트로피 모델은 전역 문맥을 사용할 수 있다.
- [33] 상기 엔트로피 부호화는 이미지 압축 네트워크 및 품질 향상 네트워크의

결합에 의해 수행될 수 있다.

- [34] 상기 품질 향상 네트워크는 매우 깊은 슈퍼 레졸루션(Very Deep Super Resolution; VDSR), 잔차 밀도 네트워크(Residual Dense Network; RDN) 또는 그룹된 잔차 밀도 네트워크(Grouped Residual Dense Network; GRDN)일 수 있다.
- [35] 상기 재구성된 이미지로부터 수평 방향의 패딩 영역 또는 수직 방향의 패딩 영역이 제거될 수 있다.
- [36] 상기 수평 방향의 패딩 영역의 제거는 상기 재구성된 이미지의 수직 축 상의 중심에서 하나 이상의 행들을 제거하는 것일 수 있다.
- [37] 상기 수직 방향의 패딩 영역의 제거는 상기 재구성된 이미지의 수평 축 상의 중심에서 하나 이상의 열들을 제거하는 것일 수 있다.
- [38] 상기 수평 방향의 패딩 영역의 제거는 원 이미지의 높이가  $k$ 의 배수가 아닐 경우에 수행될 수 있다.
- [39] 상기 수직 방향의 패딩 영역의 제거는 상기 원 이미지의 폭이  $k$ 의 배수가 아닐 경우에 수행될 수 있다.
- [40] 상기  $k$ 는  $2^n$ 일 수 있다.
- [41] 상기  $n$ 은 상기 원 이미지에 대한 다운-스케일링들의 개수일 수 있다.

### 발명의 효과

- [42] 전역적 문맥을 이용하는 기계 학습에 기반하여 이미지에 대한 압축을 제공하는 부호화 장치, 부호화 방법, 복호화 장치 및 복호화 방법이 제공된다.

### 도면의 간단한 설명

- [43] 도 1은 일 예에 따른 엔트로피 모델에 기반하는 엔드-투-엔드 이미지 압축을 나타낸다.
- [44] 도 2는 일 예에 따른 자동회귀 접근방식으로서의 확장을 나타낸다.
- [45] 도 3은 일 실시예에 따른 자동 부호기의 구현을 나타낸다.
- [46] 도 4는 일 예에 따른 이미지에 대한 훈련가능한 변수들을 나타낸다.
- [47] 도 5는 클립된 상대적 위치들을 사용하는 유도를 나타낸다.
- [48] 도 6은 일 예에 따른 (0, 0)의 현재 위치에 대한 오프셋을 도시한다.
- [49] 도 7은 일 예에 따른 (2, 3)의 현재 위치에 대한 오프셋을 도시한다.
- [50] 도 8는 일 실시예에 따른 캐스케이드로된 이미지 압축 및 품질 향상의 엔드-투-엔드 조인트 학습 스킴을 나타낸다.
- [51] 도 9는 일 실시예에 따른 이미지 압축 네트워크의 전반적인 네트워크 아키텍처를 나타낸다.
- [52] 도 10은 일 예에 따른 모델 파라미터 추정자의 구조를 나타낼 수 있다.
- [53] 도 11은 일 예에 따른 비-로컬 문맥 프로세싱 네트워크를 나타낼 수 있다.
- [54] 도 12는 일 예에 따른 오프셋-문맥 프로세싱 네트워크를 나타낼 수 있다.
- [55] 도 13은 일 예에 따른 전역 문맥 지역에 매핑된 변수들을 나타낸다.
- [56] 도 14는 일 실시예에 따른 GRDN의 구조를 나타낸다.

- [57] 도 15는 일 실시예에 따른 GRDN의 GRDB의 구조를 나타낸다.
- [58] 도 16은 일 실시예에 따른 GRDB의 RDB의 구조를 나타낸다.
- [59] 도 17은 실시예에 따른 부호기를 나타낸다.
- [60] 도 18은 실시예에 따른 복호기를 나타낸다.
- [61] 도 19는 일 실시예에 따른 부호화 장치의 구조도이다.
- [62] 도 20은 일 실시예에 따른 복호화 장치의 구조도이다.
- [63] 도 21는 일 실시예에 따른 부호화 방법의 흐름도이다.
- [64] 도 22는 일 실시예에 따른 복호화 방법의 흐름도이다.
- [65] 도 23은 일 예에 따른 입력 이미지로의 패딩을 나타낸다.
- [66] 도 24는 일 실시예에 따른 부호화에서의 패딩을 위한 코드를 나타낸다.
- [67] 도 25는 일 실시예에 따른 부호화에서의 패딩 방법의 흐름도이다.
- [68] 도 26은 일 실시예에 따른 부호화에서의 패딩 영역의 제거를 위한 코드를 나타낸다.
- [69] 도 26은 일 실시예에 따른 부호화에서의 패딩의 제거 방법의 흐름도이다.

### 발명의 실시를 위한 형태

- [70] 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 발명의 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.
- [71] 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다. 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.
- [72] 실시예에서 사용된 용어는 실시예를 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 실시예에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소, 단계, 동작 및/또는 소자는 하나 이상의 다른 구성요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않으며, 추가적인 구성이 예시적 실시예들의 실시 또는 예시적 실시예들의 기술적 사상의 범위에 포함될 수 있음을 의미한다. 어떤 구성요소(component)가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기의

2개의 구성요소들이 서로 간에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있으나, 상기의 2개의 구성요소들의 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.

- [73] 제1 및 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기의 구성요소들은 상기의 용어들에 의해 한정되어서는 안 된다. 상기의 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하여 지칭하기 위해서 사용된다. 예를 들어, 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.
- [74] 또한, 실시예들에 나타나는 구성요소들은 서로 다른 특징적인 기능들을 나타내기 위해 독립적으로 도시되는 것으로, 각 구성요소가 분리된 하드웨어나 하나의 소프트웨어 구성 단위로만 이루어짐을 의미하지 않는다. 즉, 각 구성요소는 설명의 편의상 각각의 구성요소로 나열된 것이다. 예를 들면, 구성요소들 중 적어도 두 개의 구성요소들이 하나의 구성요소로 합쳐질 수 있다. 또한, 하나의 구성요소가 복수의 구성요소들로 나뉠 수 있다. 이러한 각 구성요소의 통합된 실시예 및 분리된 실시예 또한 본질에서 벗어나지 않는 한 권리범위에 포함된다.
- [75] 또한, 일부의 구성요소는 본질적인 기능을 수행하는 필수적인 구성요소는 아니고 단지 성능을 향상시키기 위한 선택적 구성요소일 수 있다. 실시예들은 실시예의 본질을 구현하는데 필수적인 구성부만을 포함하여 구현될 수 있고, 예를 들면, 단지 성능 향상을 위해 사용되는 구성요소와 같은, 선택적 구성요소가 제외된 구조 또한 권리 범위에 포함된다.
- [76] 이하에서는, 기술분야에서 통상의 지식을 가진 자가 실시예들을 용이하게 실시할 수 있도록 하기 위하여, 첨부된 도면을 참조하여 실시예들을 상세히 설명하기로 한다. 실시예들을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 명세서의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.
- [77] 명세서의 설명에서, 기호 "/"는 "및/또는"의 약자로서 사용될 수 있다. 말하자면, "A/B"는 "A 및/또는 B"나 "A 및 B 중 적어도 하나"를 의미할 수 있다.
- [78] 전역적 문맥을 이용하는 기계 학습 기반의 이미지 압축
- [79] 최근, 인공 신경망에서의 상당한 발전은 인해 다양한 연구 분야에서 다수의 획기적인 성과들을 이끌어 왔다. 이미지 및 비디오 압축 분야에서, 다수의 학습 기반 연구들이 수행되었다.
- [80] 특히, 엔트로피 최소화에 기반하는 몇몇 최신의 엔드-투-엔드(end-to-end) 최적화 이미지 압축 방법은 이미 BPG 및 JPEG2000과 같은 기존의 이미지 압축 코덱보다 더 나은 압축 성능을 보일 수 있다.
- [81] 현장의 짧은 역사에도 불구하고, 엔트로피 최소화를 위한 기본적인 접근방식은 분석 변환 네트워크(analysis transform network)(말하자면, 부호기(encoder)) 및

- 합성 변환 네트워크(synthesis transform network)를 훈련시켜서 분석 변환 네트워크 및 합성 변환 네트워크가 재구성된(reconstructed) 이미지들의 품질을 가능한 원본들에 가깝게 유지시키면서 변환된 은닉 표현성분들(transformed latent representations)의 엔트로피를 감소시킬 수 있게 할 수 있다.
- [82] 엔트로피 최소화 접근방식은 2 개의 다른 측면에서 보일 수 있다: 사전 확률 모델링(prior probability modeling) 및 문맥 활용(context exploitation)
- [83] 사전 확률 모델링은 엔트로피 최소화의 주 요소이며, 엔트로피 모델이 은닉 표현성분들의 실제의 엔트로피를 근사(approximate)하게 할 수 있다. 사전 확률 모델링은 훈련 및 실제의 엔트로피 디코딩 및/또는 인코딩에 대하여 키 역할(key role)을 수행할 수 있다.
- [84] 각 변환된 표현성분들에 대하여, 이미지 압축 방법은 이전에 복호화된 이웃 표현성분들 또는 몇몇 비트-할당된(bit-allocated) 부(side) 정보들과 같은 문맥(context)에 기반하여 사전 확률 모델의 파라미터들을 추산할 수 있다.
- [85] 더 나은 문맥은 모델 파라미터 추정기(model parameter estimator)에 주어진 정보로 간주될 수 있다. 이러한 정보는 은닉 표현성분들의 분포들을 더 정확하게 예측하는데 도움이 될 수 있다.
- [86] 인공 신경 네트워크들(Artificial Neural Networks; ANN)-기반 이미지 압축
- [87] 도 1은 일 예에 따른 엔트로피 모델에 기반하는 엔드-투-엔드 이미지 압축을 나타낸다.
- [88] ANN-기반 이미지 압축과 관련하여 제안된 방법들은 두 개의 스트림들로 나뉠 수 있다.
- [89] 첫 번째로, 생성적(generative) 모델들의 성공의 귀결로서, 인지적(perceptual) 품질을 타겟팅하는 몇몇 이미지 압축 접근방식들이 제안되어 왔다.
- [90] 이러한 접근방식들의 기본적인 아이디어는 자연 이미지들의 분포의 학습에 있어서, 텍스처들과 같은, 재구성된 이미지의 구조 또는 인지 품질에 큰 영향을 미치지 않는 이미지 구성요소들(components)의 생성을 허용함으로써, 심각한 인지적 손실(loss) 없이 매우 높은 압축을 가능하게 하는 것이다.
- [91] 그러나, 이러한 접근방식에 의해 생성된 이미지들이 매우 사실적임에도 불구하고, 기계-생성된(machine-created) 이미지 구성요소들의 수용가능성(acceptability)은 결국에는 다소 어플리케이션-의존적(application-dependent)일 수 있다.
- [92] 한편, 두 번째로, 생성 모델들을 사용하지 않고, 엔드-투-엔드(end-to-end) 최적화된 ANN-기반 접근방식들이 사용될 수 있다.
- [93] 이러한 접근방식에서는, 예측(prediction), 변환(transform) 및 양자화(quantization)와 같은 개별적인 도구들로 구성된 전통적인 코덱들과는 다르게, 엔드-투-엔드 최적화를 통해 전체 기능들을 커버하는 포괄적(comprehensive) 솔루션이 제공될 수 있다.
- [94] 예를 들면, 한 접근방식은 모든 단계들에서 압축된 정보를 포함하기 위해

- 이진(binary) 은닉(latent) 표현성분들(representations)의 소량을 활용할 수 있다. 각 단계는 점진적으로 품질을 향상시키는 것을 달성하기 위해 추가의 은닉 표현성분들을 더욱 더 쌓을 수 있다.
- [95] 다른 접근방식은, 전술된 접근방식의 네트워크 구조를 향상시켜서 압축 성능을 향상시킬 수 있다.
- [96] 이러한 접근방식들은 하나의 훈련된 네트워크를 통한 품질 제어에 적합한 새로운 프레임워크들을 제공할 수 있다. 이러한 접근방식들에 있어서, 반복(iteration) 단계들의 개수의 증가는 몇몇 어플리케이션들에는 부담이 될 수 있다.
- [97] 이러한 접근방식들은 최대한 높은 엔트로피를 갖는 이진 표현성분들을 추출할 수 있다. 반면, 다른 접근 방식들은 이미지 압축 문제를 어떻게 가능한 낮은 엔트로피를 갖는 이산 은닉 표현성분들(discrete latent representations)을 어떻게 검출하는(retrieve) 가로 간주한다.
- [98] 다시 말하면, 전자의 접근방식들의 목표 문제는 어떻게 고정된 개수의 표현성분 내에 가능한 많은 정보를 포함시키는가로 간주될 수 있고, 반면 후자의 접근방식들의 목표 문제는 단지 표현성분들이 충분한 개수가 주어졌을 때 어떻게 예상되는 비트-레이트를 감소시킬 수 있는가로 간주될 수 있다. 여기에서, 낮은 엔트로피는 엔트로피 코딩에 의한 낮은 비트-레이트에 대응한다고 가정될 수 있다.
- [99] 후자의 접근방식들의 목표 문제를 해결하기 위해, 접근방식들은 이산 은닉 표현성분들의 실제의 분포를 근사하기 위한 자체의 엔트로피 모델들을 채용할 수 있다.
- [100] 예를 들면, 일부 접근방식들은 엔트로피 모델들을 활용하는 새로운 프레임워크들을 제안할 수 있고, 엔트로피 모델들에 의해 생성된 결과들을 JPEG2000과 같은 기존의 코덱들과 비교함으로써 엔트로피 모델들의 성능이 입증될 수 있다.
- [101] 이러한 접근방식들에 있어서, 각 표현성분이 고정된 분포를 갖는다고 가정될 수 있다. 접근방식에 대해서, 각 표현성분에 대한 분포의 스케일을 추정하는 입력-적응적(input-adaptive) 엔트로피 모델이 사용될 수 있다. 이러한 접근방식은 표현성분들의 스케일들이 인접한 영역들 내에서 함께 변한다는 자연 이미지들의 특성에 기반할 수 있다.
- [102] 엔드-투-엔드 최적화 이미지 압축의 주요 요소들 중 하나는 은닉 표현성분들을 위한 훈련가능한 엔트로피 모델일 수 있다.
- [103] 은닉 표현성분들의 실제의 분포들은 알려져 있지 않기 때문에, 엔트로피 모델들은 은닉 표현성분들의 분포들을 근사함으로써 은닉 표현성분들을 부호화하기 위한 추정된 비트들을 계산할 수 있다.
- [104] 도 1에서,  $x$ 는 입력 이미지를 나타낼 수 있다.  $x'$ 는 출력 이미지를 나타낼 수 있다.

- [105]  $Q$ 는 양자화(quantization)를 나타낼 수 있다.
- [106]  $\hat{y}$ 는 양자화된 은닉 표현성분을 나타낼 수 있다.
- [107] 입력 이미지  $x$ 가 은닉 표현성분  $y$ 로 변환(transform)되고, 은닉 표현성분  $y$ 가  $Q$ 에 의해 양자화된 은닉 표현성분  $\hat{y}$ 로 균일하게 양자화될 때, 단순한 엔트로피 모델은  $p_{\hat{y}}(\hat{y})$ 로 표현될 수 있다. 엔트로피 모델은  $y$ 의 근사(approximation)일 수 있다.
- [108]  $m(\hat{y})$ 는  $\hat{y}$ 의 실제의 한계(marginal) 분포를 나타낼 수 있다. 엔트로피 모델  $p_{\hat{y}}(\hat{y})$ 을 사용하는 교차(cross) 엔트로피를 통해 계산된 율 추정(rate estimation)은 아래의 수학적 식 1과 같이 표현될 수 있다.
- [109] [수식1]  

$$R = \mathbb{E}_{\hat{y} \sim m}[-\log_2 p_{\hat{y}}(\hat{y})] = H(m) + D_{KL}(m || p_{\hat{y}})$$
- [110] 율 추정은  
 $\hat{y}$   
 의 실제의 엔트로피 및 추가의 비트들로 분해될 수 있다. 말하자면, 율 추정은  
 $\hat{y}$   
 의 실제의 엔트로피 및 추가의 비트들을 포함할 수 있다.
- [111] 추가의 비트들은 실제의 분포들 및 이러한 실제의 분포들에 대한 추정들 간의 불일치(mismatch)에 기인할 수 있다.
- [112] 따라서, 훈련의 프로세스 동안 율 항(rate term)  
 $R$   
 이 감소하면, 엔트로피 모델  
 $p_{\hat{y}}(\hat{y})$   
 및 근사  
 $m(\hat{y})$   
 가 가능한 가까워질 수 있으며, 또한  
 $\hat{y}$   
 의 실제의 엔트로피가 작게 되도록 다른 파라미터들이  
 $x$   
 를  
 $y$   
 로 원활하게 변환할 수 있다.
- [113] 쿨백-라이블러(Kullback-Leibler; KL)-발산(divergence)의 관점에서,  
 $R$   
 은  
 $p_{\hat{y}}(\hat{y})$   
 가 실제의 분포  
 $m(\hat{y})$

- 와 완벽하게 매치될 때 최소화될 수 있다. 이는, 상기의 방법들의 압축 성능이 본질적으로 엔트로피 모델의 성능에 의존한다는 것을 의미할 수 있다.
- [114] 도 2는 일 예에 따른 자동회귀 접근방식으로의 확장을 나타낸다.
- [115] 자동회귀(auto-regressive) 접근방식(approach)의 3 개의 양상(aspect)들로서, 구조(structure), 문맥(context) 및 프라이어(prior)가 있을 수 있다.
- [116] 구조는 다양한 빌딩 블록들(building blocks)을 어떻게 결합(combine)하는가를 의미할 수 있다. 다양한 빌딩 블록들은, 하이퍼 파라미터(hyper parameter), 스킵 연결(skip connection), 비-선형성(non-linearity), 일반화된 분할 정규화(Generalized Divisive Normalization; GDN) 및 어텐션 레이어(attention layer) 등을 포함할 수 있다.
- [117] 문맥은 모델 추정을 위해 활용되는 것을 나타낼 수 있다. 활용의 대상은 인접한 알려진 영역(adjacent known area), 위치와 관련된 정보(positional information) 및  $z$ 로부터의 부가 정보(side information) 등을 포함할 수 있다.
- [118] 프라이어는 은닉 표현성분들의 실제의 분포(distribution)를 추정하기 위해 사용되는 분포들을 의미할 수 있다. 예를 들면, 프라이어는 제로-중간 가우시안(zero-mean Gaussian) 분포, 가우시안(Gaussian) 분포, 라플라시안(Laplacian) 분포, 가우시안 스케일 혼합(Gaussian Scale Mixture) 분포, 가우시안 혼합(Gaussian Mixture) 분포 및 논-파라메틱(Non-parametric) 분포 등을 포함할 수 있다.
- [119] 실시예에서, 성능을 향상시키기 위해, 문맥들의 2 개의 타입들을 활용하는 새로운 엔트로피 모델이 제안될 수 있다. 문맥의 2 개의 타입들은, 비트-소비(bit-consuming) 문맥 및 비트-프리(bit-free) 문맥일 수 있다. 비트-프리 문맥은 자동회귀 접근방식을 위해 사용될 수 있다.
- [120] 비트-소비 문맥 및 비트-프리 문맥은 문맥이 전송(transmission)을 위한 추가적인 비트 할당(allocation)을 요구하는지 여부에 따라 구분될 수 있다.
- [121] 이러한 문맥들을 이용하여, 제안되는 엔트로피 모델은 엔트로피 모델들의 보다 일반적인 형태를 사용하여 각 은닉 표현성분의 분포를 보다 정확하게 추정하게 할 수 있다. 또한, 제안되는 엔트로피 모델은 이러한 정확한 추정을 통해 인접한 은닉 표현성분들 간의 공간적 의존성들(spatial dependencies)을 더 효율적으로 감소시킬 수 있다.
- [122] 후술될 실시예들에 의해 아래와 같은 효과가 이루어질 수 있다.
- [123] - 문맥들의 2 개의 다른 타입들을 접목시키는(incorporate) 새로운 문맥-적응적 엔트로피 모델 프레임워크가 제공될 수 있다.
- [124] - 모델 용량(capacity) 및 문맥들의 레벨의 측면에서 실시예의 방법들의 개선(improvement) 방향들(directions)이 설명될 수 있다.
- [125] - ANN 기반 이미지 압축의 도메인에서, 최대 신호 대 잡음 비(Peak Signal-to-Noise Ratio; PSNR)의 측면에서, 널리 사용되는 기존의 이미지 코덱을

- 성능에서 증가하는 테스트 결과들이 제공될 수 있다.
- [126] 또한, 실시예들에 관하여 아래와 같은 설명들이 후술될 수 있다.
- [127] 1) 엔드-투-엔드 최적화된 이미지 압축의 키 접근방식들이 소개되고, 문맥-적응적 엔트로피 모델이 제안될 수 있다.
- [128] 2) 부호기 및 복호기 모델들이 구조가 설명될 수 있다.
- [129] 3) 실험의 셋업 및 실험의 결과가 제공될 수 있다.
- [130] 4) 실시예들의 현재의 상태 및 개선 방향들이 설명될 수 있다.
- [131] 문맥-적응적 엔트로피 모델에 기반하는 엔드-투-엔드 최적화의 엔트로피 모델들
- [132] 실시예의 엔트로피 모델들은 이산 은닉 표현성분들의 분포를 근사할 수 있다. 이러한 근사를 통해 엔트로피 모델들은 이미지 압축 성능을 향상시킬 수 있다.
- [133] 실시예의 엔트로피 모델들 중 어떤 것은 비-파라미터의(non-parametric) 모델들로 가정될 수 있고, 다른 것은 표현성분 당 6 개의 가중치가 부여되는(six weighted) 제로-평균(zero-mean) 가우시안 모델로 구성된 가우시안 스케일 혼합 모델일 수 있다.
- [134] 엔트로피 모델들의 형태들이 서로 다르다고 가정되더라도, 엔트로피 모델들은 입력 적응성에 대한 고려 없이 표현성분들의 분포들을 학습하는 것에 집중한다는 공통된 특징을 가질 수 있다. 다시 말해서, 일단 엔트로피 모델이 훈련되면, 표현성분들에 대하여 훈련된 모델들은 테스트 시간 동안 임의의 입력에 대해서 고정될 수 있다.
- [135] 반면, 특정 엔트로피 모델은 표현성분들에 대하여 입력-적응적 스케일 추정을 채용할 수 있다. 이러한 엔트로피 모델에서는, 자연 이미지들로부터의 은닉 표현성분들 스케일들은 인접한 영역 내에서 함께 움직이는 경향이 있다는 가정이 적용될 수 있다.
- [136] 이러한 중복성(redundancy)을 감소시키기 위해, 엔트로피 모델은 추가 정보의 소량을 사용할 수 있다. 추가 정보는 은닉 표현성분들의 적절한 스케일 파라미터들(예를 들면, 표준 편차들)과 같이 추정될 수 있다.
- [137] 스케일 추정 외에도, 연속적인 도메인 내의 각 표현성분에 대한 사전 확률 밀도 함수(Probability Density Function; PDF)가 표준 균일 밀도 함수(standard uniform density function)와 콘볼루션될(convolved) 때, 엔트로피 모델은 라운딩(rounding)에 의해 균일하게 양자화된 이산 은닉 표현성분의 사건의 확률 질량 함수(Probability Mass Function; PMF)에 더 가깝게 근사할 수 있다.
- [138] 훈련에 대하여, 균일 노이즈가 각 은닉 표현성분에 추가될 수 있다. 이러한 추가는 노이즈 낀(noisy) 표현성분들의 분포를 언급된 PMF-근사 함수들에 맞추기 위한 것일 수 있다.
- [139] 이러한 접근방식들로, 엔트로피 모델은 베타 포터블 그래픽스(Better Portable Graphics; BPG)와 유사한 최신의(state-of-the-art) 압축 성능을 달성할 수 있다.
- [140] 은닉 변수들의 공간적 의존성들

- [141] 은닉 표현성분들이 콘볼루션(convolution) 신경 네트워크를 통해 변환 될 때, 동일한 콘볼루션 필터들이 공간적 구역들(regions)을 걸쳐 공유되고, 자연 이미지들은 인접한 구역들 내에서 다양한 팩터들(factors)을 공통적으로 갖기 때문에 은닉 표현성분들은 본질적으로 공간적 의존성들을 포함할 수 있다.
- [142] 엔트로피 모델에 있어서, 은닉 표현성분들의 표준 편차들을 입력-적응적으로 추정함으로써 이러한 공간 의존성들이 성공적으로 포착될 수 있고, 압축 성능이 향상될 수 있다.
- [143] 한 걸음 더 나아가서, 표준 편차 외에도, 문맥들을 활용하는 평균(mean) 추정을 통해 추정된 분포의 형태(form)가 일반화될 수 있다.
- [144] 예를 들면, 특정한(certain) 표현성분들이 공간적으로 인접한 영역 내에서 유사한 값을 갖는 경향이 있다고 가정하면, 모든 이웃 표현성분들이 10의 값을 가질 때, 현재의 표현성분이 10 또는 유사한 값들을 가질 가능성이 비교적 높다는 것이 직관적으로 추측될 수 있다. 따라서, 이러한 간단한 추정은 엔트로피를 감소시킬 수 있다.
- [145] 마찬가지로, 실시예의 방법에 따른 엔트로피 모델은 각 은닉 표현성분의 평균 및 표준 편차를 추정하기 위해 주어진 문맥을 사용할 수 있다.
- [146] 또는, 엔트로피 모델은 각 이진 표현성분의 확률을 추정함으로써 문맥-적응적 엔트로피 코딩을 수행할 수 있다.
- [147] 그러나, 이러한 문맥-적응적 엔트로피 코딩은, 엔트로피 코딩의 확률 추정이 율-왜곡(Rate-Distortion; R-D) 최적화 프레임워크의 율 항(rate term)에 직접적으로 기여하지 않기 때문에, 엔드-투-엔드 최적화 구성요소들 중 하나라기 보다는 별개의 구성요소들로 보일 수 있다.
- [148] 2 개의 상이한 접근방식들의 은닉 변수들  $m(\hat{y})$  및 이러한 은닉 변수들의 정규화된 버전들이 예시될 수 있다. 앞서 언급된 문맥들의 2 개의 타입들을 가지고, 하나의 접근방식에서는 단지 표준 편차 파라미터들이 추정될 수 있고, 다른 하나의 접근방식에서는 평균 및 표준 편차 파라미터들의 양자가 추정될 수 있다. 이 때, 주어진 문맥들을 가지고 평균이 함께 추정될 때 공간적 의존성은 더 효율적으로 제거될 수 있다.
- [149] 문맥-적응적 엔트로피 모델
- [150] 실시예에서의 최적화 문제에 있어서, 입력 이미지  $x$  는 낮은 엔트로피를 갖는 은닉 표현성분  $y$  로 변환될 수 있고,  $y$  의 공간적 의존성들은  $\hat{z}$

로 포착될 수 있다. 따라서, 4 개의 주요한 파라미터의(parametric) 변환 함수들이 사용될 수 있다. 엔트로피 모델의 4 개의 파라미터의 변환 함수들은 아래의 1) 내지 4)와 같다.

[151] 1)

$x$

를 은닉 표현성분

$y$

로 변환하기 위한 분석 변환

$g_a(\mathbf{x}; \phi_g)$

[152] 2) 재구축된 이미지

$\hat{x}$

를 생성하기 위한 합성(synthesis) 변환

$g_s(\hat{y}; \theta_g)$

[153] 2)

$\hat{y}$

의 공간적 중복성들을 은닉 표현성분

$z$

로 포착(capture)하기 위한 분석 변환

$h_a(\hat{y}; \phi_h)$

[154] 4) 모델 추정에 대한 문맥들을 생성하기 위한 합성 변환

$h_s(\hat{z}; \theta_h)$

[155] 실시예에서,

$h_s$

는 표현성분들의 표준 편차들을 직접적으로 추정하지 않을 수 있다. 대신, 실시예에서,

$h_s$

는 분포를 추정하기 위해 문맥들의 복수의 개의 타입들 중 하나인 문맥

$c'$

을 생성할 수 있다. 문맥들의 복수의 개의 타입들에 대해서는 아래에서 설명된다.

[156] 변이(variational) 자동 부호기(autoencoder)의 시점(viewpoint)로부터 최적화 문제가 분석될 수 있고, KL-발산의 최소화는 이미지 압축의 R-D 최적화와 동일한 문제로 간주될 수 있다. 기본적으로, 실시예에서는 동일한 컨셉이 채용될 수 있다. 그러나 훈련에 있어서, 실시예에서는 노이즈 낀 표현성분들 대신에 조건들(conditions)에 대한 이산 표현성분들이 사용될 수 있고, 따라서 노이즈 낀 표현성분들은 엔트로피 모델들로의 입력들로만 사용될 수 있다.

[157] 경험적으로, 조건들에 대한 이산 표현성분들을 사용하는 것은 더 나은

결과들을 낳을 수 있다. 이러한 결과들은 훈련 시간 및 테스트 시간 사이에서의 조건들의 불일치를 제거하는 것과, 이러한 불일치의 제거에 의해 훈련 용량을 향상시키는 것으로부터 올 수 있다. 훈련 용량은 균일 노이즈의 영향(affect)을 단지 확률 질량 함수들로의 근사를 돕는 것만으로 제한함으로써 향상될 수 있다.

[158] 실시예에서, 균일 양자화로부터의 불연속성들(discontinuities)을 다루기 위해 정체(identity) 함수를 갖는 그래디언트 오버라이딩(gradient overriding) 방법이 사용될 수 있다. 실시예에서 사용되는 결과인(resulting) 목적 함수들(objective functions)은 아래의 수학적식 2에서 설명되었다.

[159] [수식2]

$$\mathcal{L} = R + \lambda D$$

$$\text{with } R = \mathbb{E}_{\mathbf{x} \sim p_{\mathbf{x}}} \mathbb{E}_{\tilde{\mathbf{y}}, \tilde{\mathbf{z}} \sim q} \left[ -\log p_{\tilde{\mathbf{y}}|\tilde{\mathbf{z}}}(\tilde{\mathbf{y}} | \tilde{\mathbf{z}}) - \log p_{\tilde{\mathbf{z}}}(\tilde{\mathbf{z}}) \right]$$

$$D = \mathbb{E}_{\mathbf{x} \sim p_{\mathbf{x}}} \left[ -\log p_{\mathbf{x}|\hat{\mathbf{y}}}(\mathbf{x} | \hat{\mathbf{y}}) \right]$$

[160] 수학적식 2에서, 총 손실(total loss)은 2 개의 항들을 포함한다. 2 개의 항들은 비율들 및 왜곡들을 나타낸다. 말하자면, 총 손실은 율 항(rate term)  $R$  및 왜곡 항(distortion term)  $D$ 를 포함할 수 있다.

[161] 계수

$\lambda$

는 R-D 최적화 프로세스 내에서 율 및 왜곡 간의 균형(balance)을 제어할 수 있다.

[162] [수식3]

$$q(\tilde{\mathbf{y}}, \tilde{\mathbf{z}} | \mathbf{x}, \phi_g, \phi_h) = \prod_i \mathcal{U}(\tilde{y}_i | y_i - \frac{1}{2}, y_i + \frac{1}{2}) \cdot \prod_j \mathcal{U}(\tilde{z}_j | z_j - \frac{1}{2}, z_j + \frac{1}{2})$$

$$\text{with } \mathbf{y} = g_a(\mathbf{x}; \phi_g), \hat{\mathbf{y}} = Q(\mathbf{y}), \mathbf{z} = h_a(\hat{\mathbf{y}}; \phi_h).$$

[163] 여기에서,

$\mathbf{y}$

가 변환

$g_a$

의 결과이고,

$\mathbf{z}$

가 변환

$h_a$

의 결과일 때,

$\tilde{\mathbf{y}}$

및

$\tilde{\mathbf{z}}$

의 노이즈가 긴 표현성분은 표준 균일 분포를 따를 수 있다. 여기에서,

$\tilde{y}$ 

의 평균은

 $y$ 

일 수 있고,

 $\tilde{z}$ 

의 평균은

 $z$ 

일 수 있다. 또한,

 $h_a$ 

로의 입력은, 노이즈 낀 표현성분

 $\tilde{y}$ 

가 아니라,

 $\hat{y}$ 

일 수 있다.

 $\hat{y}$ 

는 라운딩 함수

 $Q$ 

에 의한

 $y$ 

의 균일하게 양자화된 표현성분들일 수 있다.

[164] 을 향은

 $p_{\tilde{y}|\tilde{z}}$ 

및

 $p_{\tilde{z}}$ 

의 엔트로피 모델들을 가지고 계산된 예상되는 비트들을 나타낼 수 있다.

 $p_{\tilde{y}|\tilde{z}}$ 

는 궁극적으로

 $p_{\hat{y}|\tilde{z}}$ 

의 근사일 수 있고,

 $p_{\tilde{z}}$ 

는 궁극적으로

 $p_{\tilde{z}}$ 

의 근사일 수 있다.

[165] 아래의 수학식 4는

 $\hat{y}$ 

에 대한 요구되는 비트들의 근사를 위한 엔트로피 모델을 나타낼 수 있다.

수학식 4는 엔트로피 모델에 대한 공식적인(formal) 표현성분일 수 있다.

[166] [수식4]

$$p_{\tilde{\mathbf{y}}|\hat{\mathbf{z}}}(\tilde{\mathbf{y}} | \hat{\mathbf{z}}, \boldsymbol{\theta}_h) = \prod_i \left( \mathcal{N}(\mu_i, \sigma_i^2) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right) (\tilde{y}_i)$$

with  $\mu_i, \sigma_i = f(\mathbf{c}'_i, \mathbf{c}''_i)$ ,  
 $\mathbf{c}'_i = E'(h_s(\hat{\mathbf{z}}; \boldsymbol{\theta}_h), i)$ ,  
 $\mathbf{c}''_i = E''(\langle \hat{\mathbf{y}} \rangle, i)$ ,  
 $\hat{\mathbf{z}} = Q(\mathbf{z})$

[167] 엔트로피 모델은 표준 편차 파라미터

$\sigma_i$

뿐만 아니라, 평균 파라미터

$\mu_i$

도 갖는 가우시안 모델에 기반할 수 있다.

[168]

$\sigma_i$

및

$\mu_i$

는 함수

$f$

에 의해 주어진 문맥들의 2 개의 타입들로부터 결정적 방식으로 추정될 수 있다. 함수

$f$

는 추정자(estimator)일 수 있다. 실시예에서, 용어들 "추정자", "분포 추정자", "모델 추정자" 및 "모델 파라미터 추정자"는 동일한 의미를 가질 수 있으며, 서로 교체되어 사용될 수 있다.

[169] 문맥들의 2 개의 타입들은 비트-소비 문맥 및 비트-프리 문맥일 수 있다.

여기에서, 어떤 표현성분의 분포를 추정하기 위한 문맥들의 2 개의 타입들은

$\mathbf{c}'_i$

및

$\mathbf{c}''_i$

로 표시될 수 있다.

[170] 추출자

$E'$

는

$\mathbf{c}'$

로부터

$\mathbf{c}'_i$

를 추출할 수 있다.

- $c'$   
 는 변환  
 $h_s$   
 의 결과일 수 있다.
- [171]  $c'$   
 와는 대조적으로,  
 $c''_i$   
 에 대해서는 어떤 추가 비트 할당도 요구되지 않을 수 있다. 대신,  
 $\hat{y}$   
 의 알려진(이미 엔트로피-부호화되거나, 엔트로피-복호화된) 서브세트가  
 활용될 수 있다. 이러한  
 $\hat{y}$   
 의 알려진 서브세트는  
 $\langle \hat{y} \rangle$   
 로 표시될 수 있다.
- [172] 추출자  
 $E''$   
 는  
 $\langle \hat{y} \rangle$   
 로부터  
 $c''_i$   
 를 추출할 수 있다.
- [173] 엔트로피 부호기 및 엔트로피 복호기는, 래스트 스캐닝(raster scanning)과 같은,  
 동일한 특정된(specific) 순서로 순차적으로(sequentially)  
 $\hat{y}_i$   
 를 처리할 수 있다. 따라서, 동일한  
 $\hat{y}_i$   
 를 처리함에 있어서, 엔트로피 부호기 및 엔트로피 복호기에 주어지는  
 $\langle \hat{y} \rangle$   
 는 언제나 동일할 수 있다.
- [174]  $\hat{z}$   
 의 경우에는, 단순한 엔트로피 모델이 사용될 수 있다. 이러한 단순한 엔트로피  
 모델은 훈련가능한  
 $\sigma$   
 를 가진 제로-평균 가우시안 분포들을 따르는 것으로 가정될 수 있다.
- [175]  $\hat{z}$   
 는 부가 정보(side information)로 간주될 수 있으며,

$\tilde{z}$

는 총 비트-레이트의 매우 적은 양에 기여할 수 있다. 따라서, 실시예에서는, 더 복잡한 엔트로피 모델들이 아닌, 엔트로피 모델의 단순화된 버전이 제안된 방법의 전체의 파라미터들 상의 엔드-투-엔드 최적화를 위해 사용될 수 있다.

[176] 아래의 수학식 5는 엔트로피 모델의 단순화된 버전을 나타낸다.

[177] [수식5]

$$p_{\tilde{z}}(\tilde{\mathbf{z}}) = \prod_j \left( \mathcal{N}(0, \sigma_j^2) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right)(\tilde{z}_j)$$

[178] 율 항은 실제의 비트들의 양이 아니고, 언급된 것과 같이 엔트로피 모델들로부터 계산된 추정일 수 있다. 따라서, 훈련 또는 부호화에 있어서, 실제의 엔트로피 부호화 또는 엔트로피 복호화 프로세스들이 필수적으로 요구되지 않을 수 있다.

[179] 왜곡 항(distortion term)에 관하여,

$P_{\mathbf{x}|\hat{\mathbf{y}}}$

가 널리-사용되는 왜곡 메트릭스들(metrics)로서 가우시안 분포들을 따른다고 가정될 수 있다. 이러한 가정 하에서, 왜곡 항은 평균 제곱된 에러(Mean Squared Error; MSE)를 사용하여 계산될 수 있다.

[180] 도 3은 일 실시예에 따른 자동 부호기의 구현을 나타낸다.

[181] 도 3에서, 콘볼루션은 "conv"로 약술되었다. "GDN"은 일반화된 분할 정규화(generalized divisive normalization)를 나타낼 수 있다. "IGDN"은 역 일반화된 분할 정규화(inverse generalized divisive normalization)를 나타낼 수 있다.

[182] 도 3에서, leakyReLU는 ReLU의 변형인 함수일 수 있으며, 유출되는(leaky) 정도가 특정되는 함수일 수 있다. leakyReLU 함수에 대해 제1 설정 값 및 제2 설정 값이 설정될 수 있다. leakyReLU 함수는 입력 값이 제1 설정 값의 이하인 경우, 제1 설정 값을 출력하지 않고, 입력 값 및 제2 설정 값을 출력할 수 있다.

[183] 또한, 도 3에서 사용된 콘볼루션 레이어에 대한 기호법들(notations)은 다음과 같을 수 있다: 필터들의 개수

×

필터 높이

×

필터 폭 (/ 다운-스케일 또는 업-스케일의 팩터(factor)).

[184] 또한,

↑

및

↓

는 업-스케일링 및 다운-스케일링을 각각 나타낼 수 있다. 업-스케일링 및

다운-스케일링에 대해서, 트랜스포스된(transposed)된 컨볼루션이 사용될 수 있다.

[185] 컨볼루션 신경 네트워크들은 변환 및 재구축 기능들을 구현하기 위해 사용될 수 있다.

[186] 도 3에서 도시된

$g_a$

,

$g_s$

,

$h_a$

및

$h_s$

는 전술된 다른 실시예에서의 설명이 적용될 수 있다. 또한,

$h_s$

의 말단(end)에서는, 절대(absolute) 연산자(operator)가 아닌 자승(exponentiation) 연산자가 사용될 수 있다.

[187] 각

$\hat{y}_i$

의 분포를 추정하기 위한 구성요소들이 컨볼루션 자동 부호기에 추가되었다.

[188] 도 3에서, "Q"는 균일 양자화(반올림)을 나타낼 수 있다. "EC"는 엔트로피 인코딩을 나타낼 수 있다. "ED"는 엔트로피 디코딩을 나타낼 수 있다. "

$f$

"는 분포 추정자를 나타낼 수 있다.

[189] 또한, 컨볼루션 자동 부호기는 컨볼루션 레이어들을 사용하여 구현될 수 있다. 컨볼루션 레이어로의 입력은 채널-단위로(channel-wisely) 연쇄된(concatenated)

$c'_i$

및

$c''_i$

일 수 있다. 컨볼루션 레이어는 추정된

$\mu_i$

및 추정된

$\sigma_i$

를 결과들로서 출력할 수 있다.

[190] 여기에서, 동일한

$c'_i$

및

$c''_i$

가 동일한 공간적 위치에 위치하는 모든

$\hat{y}_i$

들에게 공유될 수 있다.

[191]  $E'$

는

$c'_i$

를 검출하기 위해 채널들을 걸쳐 모든 공간적으로 인접한 요소들을

$c'$

로부터 추출할 수 있다. 유사하게,

$E''$

는

$c''_i$

를 위하여 모든 인접한 알려진 요소들을

$\langle \hat{y} \rangle$

로부터 추출할 수 있다. 이러한

$E'$

및

$E''$

에 의한 추출들은 서로 다른 채널들 사이의 남아있는(remaining)

상관관계들(correlations)을 캡춰하는 효과를 가질 수 있다.

[192]  $f$

는 동일한 공간적 위치에서의 1) 모든

$M$

, 2)

$y$

의 채널들의 총 개수 및 3)

$\hat{y}_i$

들의 분포들을 단 하나의 단계에서 추출할 수 있으며, 이러한 추출을 통해 추정들의 총 개수가 감소될 수 있다.

[193] 나아가

$f$

의 파라미터들은

$\hat{y}$

의 모든 공간적 위치들에 대하여 공유될 수 있다. 이러한 공유를 통해

$\lambda$

당 단지 하나의 훈련된

$f$

가 이미지들의 임의의 크기를 처리하기 위해 필요할 수 있다.

[194] 그러나, 훈련의 경우, 전술된 단순화들에도 불구하고, 율 항을 계산하기 위해 전체의 공간적 위치들로부터의 결과들을 수집하는 것은 크나큰 부담이 될 수 있다. 이러한 부담을 감소시키기 위해, 문맥 적응형 엔트로피 모델에 대한 모든 훈련 단계마다 랜덤의(random) 공간적 포인트들의 특정된 개수(예를 들면, 16)가 대표자들(representatives)로서 지정될 수 있다. 이러한 지정은 율 항의 계산을 용이하게 할 수 있다. 여기에서, 이러한 랜덤 공간적 포인트들은 단지 율 항을 위해서 사용될 수 있다. 반면, 왜곡 항은 여전히 전체의 이미지들 상에서 계산될 수 있다.

[195]  $y$

는 3-차원의 배열(array)이기 때문에,

$y$

에 대한 인덱스  $i$ 는 3 개의 인덱스들  $k, l$  및  $m$ 을 포함할 수 있다.  $k$ 는 수평의 인덱스일 수 있다.  $l$ 는 수직의 인덱스일 수 있다.  $m$ 는 채널 인덱스일 수 있다.

[196] 현재의 위치가  $(k, l, m)$ 일 때,

$E'$

는

$c'_{[k-2\dots k+1],[l-3\dots l],[1\dots M]}$

을

$c'_i$

로서 추출할 수 있다. 또한,

$E''$

는

$\langle \hat{y} \rangle_{[k-2\dots k+1],[l-3\dots l],[1\dots M]}$

를

$c''_i$

로서 추출할 수 있다. 여기에서,

$\langle \hat{y} \rangle$

는

$\hat{y}$

의 알려진 영역을 나타낼 수 있다.

[197]  $\langle \hat{y} \rangle$

의 알려지지 않은 영역은 0으로 채워질 수 있다.

$\langle \hat{y} \rangle$

의 알려지지 않은 영역을 0으로 채움에 따라,

$\langle \hat{y} \rangle$

의 차원이

$\hat{y}$ 

의 차원과 동일성을 갖도록 유지될 수 있다. 따라서,

 $c''_{i \in [3...4], 4, [1...M]}$ 

는 언제나 0으로 채워질 수 있다.

[198] 추정 결과들의 차원을 입력으로 유지시키기 위해,

 $c'$ 

및

 $\langle \hat{y} \rangle$ 

의 마진의(marginal) 영역들 또한 0으로 세트될 수 있다.

[199] 훈련 또는 부호화가 수행될 때,

 $c''_i$ 

는 단지 단순한 4

×

4

×

$M$

윈도우들 및 이진(binary) 마스크들을 사용하여 추출될 수 있다. 이러한 추출은 병렬 처리를 가능하게 할 수 있다. 반면, 복호화에서는, 순차적인(sequential) 재구축이 사용될 수 있다.

[200] 구현 비용을 감소시키기 위한 다른 구현 테크닉으로서, 하이브리드 접근방식이 사용될 수 있다. 실시예의 엔트로피 모델은 경량(lightweight) 엔트로피 모델과 결합될 수 있다. 경량 엔트로피 모델에 있어서, 표현성분들은 추정된 표준편차들을 갖는 제로-평균 가우시안 모델을 따르는 것으로 가정될 수 있다.

[201] 이러한 하이브리드 접근방식은 9 개의 구성들(configurations) 내에서 비트-레이트의 내림차순으로 상위 4 개의 경우들에 대하여 활용될 수 있다. 이러한 활용에 있어서, 더 고품질의 압축에 대하여 매우 낮은 공간적 의존성을 갖는 희소(sparse) 표현성분들의 개수가 증가하고, 따라서 직접 스케일 추정이 이러한 추가된 표현성분들에 대해서 충분한 성능을 제공한다는 것이 가정될 수 있다.

[202] 구현에 있어서, 은닉 표현성분

 $y$ 

는 2 개의 파트들

 $y_1$ 

및

 $y_2$ 

로 분리될 수 있다. 2 개의 상이한 엔트로피 모델들이

 $y_1$

및

$y_2$

에 대해서 적용될 수 있다.

$g_a$

,

$g_s$

,

$h_a$

및

$h_s$

의 파라미터들은 공유될 수 있고, 전체의 파라미터들은 여전히 함께 훈련될 수 있다.

[203] 예를 들면, 5 개의 하위의 구성들에 대하여 파라미터들

$N$

의 개수는 182로 세트될 수 있다. 파라미터들

$M$

의 개수는 192로 세트될 수 있다. 약간 더 많은 파라미터들이 더 상위의 구성들에 대해서 사용될 수 있다.

[204] 실제의 엔트로피 부호화를 위해, 산술(arithmetic) 부호기가 사용될 수 있다. 산술 부호기는 추정된 모델 파라미터들을 가지고 전송된 것과 같은 비트스트림의 생성 및 재구축을 수행할 수 있다.

[205] 앞서 설명된 것과 같이, 엔트로피 모델을 활용하는 ANN-기반 이미지 압축 접근방식에 기반하여, 실시예의 엔트로피 모델들은 문맥들의 2 개의 다른 타입들을 활용하도록 확장될 수 있다.

[206] 이러한 문맥들은 엔트로피 모델이 평균 파라미터들 및 표준 편차들을 갖는 일반화된 형태를 갖고서 표현성분들의 분포를 더 정확하게 추정하게 할 수 있다.

[207] 활용되는 문맥들은 2 개의 타입들로 나뉠 수 있다. 2 개의 타입들 중 하나는 자유(free) 문맥의 일종일 수 있으며, 부호기 및 복호기의 양자에게 알려진 은닉 변수들의 부분을 포함할 수 있다. 2 개의 타입들 중 다른 하나는 공유될 추가의 비트의 할당을 요하는 문맥일 수 있다. 전자는 다양한 코덱들에서 일반적으로 이용되는 문맥들일 수 있다. 후자는 압축에 도움이 되는 것으로 검증된 것일 수 있다. 실시예에서는, 이러한 문맥들을 활용하는 엔트로피 모델들의 프레임워크가 제공되었다.

[208] 추가적으로 실시예의 성능을 향상시키는 다양한 방법들이 고려될 수 있다.

[209] 성능 향상을 위한 하나의 방법은 엔트로피 모델의 기반이 되는 분포 모델을 일반화하는 것일 수 있다. 실시예에서는, 이전의 엔트로피 모델들을 일반화함으로써 성능이 향상될 수 있고, 상당히 수용 가능한 결과가 검출될 수 있다. 그러나, 가우시안-기반의 엔트로피 모델들은 명백하게 제한된

표현력(expression power)을 가질 수 있다.

- [210] 예를 들면, 비-파라미터의(non-parametric) 모델들과 같이 더 정교한(elaborate) 모델들이 실시예의 문맥-적응성(context-adaptivity)과 결합될 경우, 이러한 결합은 실제의 분포들 및 추정 모델들 간의 미스매치를 감소시킴으로써 더 나은 결과들을 제공할 수 있다.
- [211] 성능 향상을 위한 다른 방법은 문맥들의 레벨들을 향상시키는 것일 수 있다.
- [212] 실시예는 제한된 인접 영역들 내에서의 낮은 레벨의 표현성분들을 사용할 수 있다. 네트워크들의 충분한 용량과, 문맥들의 더 높은 레벨이 주어진다면, 실시예에 의해 더 정확한 추정이 가능해질 수 있다.
- [213] 예를 들면, 사람 안면들의 구조들에 관하여, 엔트로피 모델이 상기의 구조들이 일반적으로 2 개의 눈들을 가지고, 2 개의 눈들 간의 대칭이 존재한다는 것을 이해한다면, 엔트로피 모델은 사람 안면의 남은 하나의 눈을 부호화함에 있어서 (하나의 눈의 형상과 위치를 참조하여) 분포들을 더 정확하게 근사할 수 있다.
- [214] 예를 들면, 생성적인 엔트로피 모델은, 예를 들면 사람 안면들 및 침실들과 같은 특정한 도메인 내에서의 이미지들의 분포  $p(\mathbf{x})$  를 학습할 수 있다. 또한, 인-페인팅(in-painting) 방법들은 보이는 영역들이 *context* 로 주어졌을 때 조건적인(conditional) 분포  $p(\mathbf{x} | \text{context})$  를 학습할 수 있다. 이러한 고-레벨 이해들(understandings)이 실시예에 결합될 수 있다.
- [215] 나아가, 부가 정보를 통해 제공되는 문맥들은 세그멘테이션 맵(segmentation map) 및 압축을 돕는 다른 정보과 같은 고-레벨 정보로 확장될 수 있다. 예를 들면, 세그멘테이션 맵은 표현성분이 속하는 세그먼트 클래스에 따라 표현성분의 분포를 구별적으로(discriminatively) 추정하는 것을 도울 수 있다.
- [216] 향상된 엔트로피 최소화를 갖는 이미지 압축 및 품질 향상의 엔드-투-엔드 조인트 학습 스킴
- [217] 실시예의 엔드-투-엔드 조인트 학습 스킴과 관련하여 아래의 기술이 사용될 수 있다:
- [218] 1) 엔트로피 모델에 기반한 접근방식들: 엔드-투-엔드 최적화 이미지 압축이 사용될 수 있으며, 압축적인(compressive) 자동부호기를 갖는 손실(lossy) 이미지 압축이 사용될 수 있다.
- [219] 2) 은닉 표현성분들의 계층적인(hierarchical) 프라이어(prior) 추정 스케일 파라미터들: 스케일 하이퍼프라이어(hyperprior)를 갖는 변하는(variational) 이미지 압축이 사용될 수 있다.
- [220] 3) 하이퍼프라이어로부터의 문맥과 조인트하여 인접한(adjacent) 은닉

표현성분들을 추가의 문맥으로 활용: 조인트 자동회귀(autoregressive) 및 계층적 프라이어를 학습된 이미지 압축을 위해서 사용될 수 있고, 엔드-투-엔드 최적화된 이미지 압축에 대하여 문맥-적응적(context-adaptive) 엔트로피 모델이 사용될 수 있다.

- [221] 실시예에서, 문맥에 대해서 아래와 같은 특성이 고려될 수 있다:
- [222] 1) 공간적 상관관계(spatial correlation): 자동회귀에 있어서 기존의 접근방법들은 단지 인접한 영역들만 활용할 수 있다. 그러나, 많은 표현성분들은 실-세계(real-image) 이미지 내에서 반복될 수 있다. 남아있는(remaining) 비-로컬 상관관계들은 제거될 필요가 있다.
- [223] 2) 채널-간(inter-channel) 상관관계: 은닉 표현성분들 내의 서로 다른 채널들 간의 상관관계는 효율적으로 제거될 수 있다. 또한, 채널-간 상관관계가 활용될 수 있다.
- [224] 따라서, 문맥에 대하여, 실시예에서는 새롭게 정의된 비-로컬 문맥과의 공간적 상관관계가 제거될 수 있다.
- [225] 실시예에서, 구조에 대해서 아래와 같은 특성이 고려될 수 있다: 품질 향상을 위한 방법들이 이미지 압축에 조인트되어 최적화될 수 있다.
- [226] 실시예에서, 프라이어에 대해서 아래와 같은 문제 및 특성이 고려될 수 있다: 가우시안 프라이어를 사용하는 접근방법은 표현력(expression power)에 제한을 가질 수 있으며, 실제의 분포들에 피팅함에 있어서 제약을 가질 수 있다. 프라이어가 더 일반화될(generalized)수록, 실제의 분포들에 대한 더 정확한 근사(approximation)을 통해, 더 높은 압축 성능이 획득될 수 있다.
- [227] 도 4는 일 예에 따른 이미지에 대한 훈련가능한 변수들을 나타낸다.
- [228] 도 5는 클립된 상대적 위치들을 사용하는 유도를 나타낸다.
- [229] 비-로컬 상관관계들을 제거하는 문맥에 대해서 아래의 요소들이 사용될 수 있다:
- [230] - 각 채널에 대한, 알려진 은닉 표현성분들의 가중치가 부여된 샘플 평균(average) 및 변화(variance)
- [231] - 가변-크기 영역들에 대한 고정된 가중치들
- [232] 비-로컬 문맥은 비-로컬 상관관계들을 제거하는 문맥을 의미할 수 있다.
- [233] 비-로컬 문맥

$$c_i^{n,l}$$

은 아래의 수학식 6과 같이 정의될 수 있다.

[234] [수식6]

$$c_i^{n.l.} = \{\mu_0^*, \dots, \mu_J^*, \sigma_0^*, \dots, \sigma_J^*\}$$

$$\text{with } \sigma_j^* = \sqrt{\frac{\sum_{k,l \in S} w_{j,k,l} (h_{j,k,l} - \mu_j^*)^2}{1 - \sum_{k,l \in S} w_{j,k,l}^2}},$$

$$\mu_j^* = \sum_{k,l \in S} w_{j,k,l} h_{j,k,l},$$

[235] 수학식 6에 대하여, 아래의 수학식 7 및 수학식 8이 사용될 수 있다.

[236] [수식7]

$$h = H(\langle \hat{y} \rangle),$$

$$\text{with } \langle \hat{y} \rangle = \{y_{j,k,l} | k, l \in S\}$$

[237] [수식8]

$$w = \{w_0, \dots, w_J\}$$

$$\text{with } w_j = \text{softmax}(a_j),$$

$$a_j = \{a_{j,k,l} | k, l \in S\},$$

$$a_{j,k,l} = v_{j, \text{clip}(k-k_{\text{cur}}, K), \text{clip}(l-l_{\text{cur}}, K)},$$

$$\text{clip}(x, K) = \max(-K, \min(K, x))$$

[238]  $H$ 는 선형 함수를 나타낼 수 있다.

[239]  $j$ 는 채널에 대한 인덱스일 수 있다.  $k$ 는 수직 축에 대한 인덱스일 수 있다.  $l$ 은 수평 축에 대한 인덱스일 수 있다.

[240]  $k$ 는  $v_j$  내의 훈련가능한 변수들의 개수를 결정하는 항수(constant)일 수 있다.

[241] 도 4에서는, 현재 위치에 대한 훈련가능한 변수들  $v_j$ 이 도시되었다.

[242] 현재 위치는 부호화 및/또는 복호화의 대상의 위치일 수 있다.

[243] 훈련가능한 변수들은 현재 위치로부터의 거리가  $k$ 의 이하인 변수들일 수 있다. 현재 위치로부터의 거리는 1) 현재의  $x$  좌표 및 변수의  $x$  좌표 간의 차이 및 2) 현재의  $y$  좌표 및 변수의  $y$  좌표 간의 차이 중 더 큰 것일 수 있다.

[244] 도 5에서는, 클립된 상대적 위치들(clipped relative positions)을 사용하여 유도된 변수들이 도시되었다.

[245] 도 5에서, 현재 위치는 (9, 11)이고, 폭은 13인 것으로 예시되었다.

[246] 도 6은 일 예에 따른 (0, 0)의 현재 위치에 대한 오프셋을 도시한다.

[247] 도 7은 일 예에 따른 (2, 3)의 현재 위치에 대한 오프셋을 도시한다.

[248] 실시예에서, 경계들(boarders)로부터의 오프셋들을 가리키는 문맥이 사용될 수 있다.

[249] 마진 영역들 내의 제로-값들로부터의 모호성(ambiguity) 때문에, 은닉

- 표현성분들의 조건적인(conditional) 분배들은 공간적(spatial) 위치들에 따라서 다를 수 있다. 이러한 특징을 고려하여, 오프셋들이 문맥들로서 활용될 수 있다.
- [250] 오프셋은 경계들로부터의 오프셋들을 가리키는 문맥을 의미할 수 있다.
- [251] 도 6 및 도 7에서는 현재 위치, 유효 영역(effective area) 및 마진 영역(margin area)이 도시되었다.
- [252] 도 6에서, 오프셋( $L, R, T, B$ )는  $(0, w-1, 0, h-1)$ 일 수 있고, 도 7에서, 오프셋( $L, R, T, B$ )는  $(2, w-3, 3, h-4)$ 일 수 있다.
- [253]  $L, R, T$  및  $B$ 는 각각 좌측, 우측, 상단 및 하단을 의미할 수 있다.  $w$ 는 입력 이미지의 폭일 수 있다.  $h$ 는 입력 이미지의 높이일 수 있다.
- [254] 네트워크 아키텍처
- [255] 이미지 압축 및 품질 향상의 조인트 학습 스킴(joint learning scheme)
- [256] 도 8는 일 실시예에 따른 캐스케이드로된 이미지 압축 및 품질 향상의 엔드-투-엔드 조인트 학습 스킴을 나타낸다.
- [257] 도 8에서는, 품질 향상 네트워크들을 포용(embracing)하는 구조(structure)들이 도시되었다.
- [258] 실시예에서, 개시되는 이미지 압축 네트워크는 엔드-투-엔드 조인트 학습 스킴에 대하여 기존의 이미지 품질 향상 네트워크를 채용할 수 있다. 이미지 압축 네트워크는 이미지 압축 및 품질 향상을 조인트로 최적화할 수 있다.
- [259] 따라서, 실시예의 아키텍처는 고 유연성(flexibility) 및 고 확장성(extensibility)을 제공할 수 있다. 특히, 실시예의 방법은 미래의 향상된 이미지 품질 향상 네트워크들을 용이하게 수용(accommodate)할 수 있으며, 이미지 압축 방법들 및 품질 향상 방법들의 다양한 조합들을 허용할 수 있다. 즉, 개별적으로 개발된 이미지 압축 네트워크들 및 이미지 향상 네트워크들은 아래의 수학식 9의 총 손실(total loss)를 최소화하는 통합된 아키텍처 내에서 용이하게 조합될 수 있고, 용이하게 조인트되어 최적화될 수 있다.
- [260] [수식9]
- $$\mathcal{L} = R + \lambda D(x, Q(I(x)))$$
- [261]  $\mathcal{L}$   
는 총 손실을 나타낼 수 있다.
- [262]  $I$   
는 입력 이미지  
 $x$   
를 입력으로 사용하는 이미지 압축을 나타낼 수 있다. 말하자면,  
 $I$   
는 이미지 압축 서브-네트워크일 수 있다.
- [263]  $Q$   
는 재구성된 이미지

$\hat{x}$

를 입력으로 사용하는 품질 향상 함수일 수 있다. 말하자면,

$Q$

는 품질 향상 서브-네트워크일 수 있다.

[264] 여기서,

$\hat{x}$

는

$I(\mathbf{x})$

일 수 있다. 또한,

$\hat{x}$

는

$I$

,

$R$

,

$D$

및

$\lambda$

의 중간(intermediate) 재구축 출력일 수 있다.

[265]  $R$

는 율(rate)을 나타낼 수 있다.

[266]  $D$

는 왜곡(distortion)을 나타낼 수 있다.

$D(x, Q(I(\mathbf{x})))$

는

$\mathbf{x}$

및

$Q(I(\mathbf{x}))$

간의 왜곡을 나타낼 수 있다.

[267]  $\lambda$

는 균형 파라미터(balancing parameter)를 나타낼 수 있다.

[268] 종래의 방법들에서는 이미지 압축 서브-네트워크

$I$

를 출력 이미지들이 가능한 작은 왜곡을 갖도록 재구축하도록 훈련시킬 수 있다. 이러한 종래의 방법들과 대비되게, 실시예에서

$I$

의 출력들은 중간 은닉 표현성분

$\hat{x}$ 

으로 간주될 수 있다.

 $\hat{x}$ 

은 품질 향상 서브-네트워크

 $Q$ 

로 입력될 수 있다.

[269] 따라서, 왜곡

 $D$ 

는 1) 입력 이미지

 $x$ 

및 2)

 $Q$ 

에 의해 재구성된 최종의 출력 이미지

 $x'$ 

의 사이에서 측정될 수 있다.

[270] 여기에서,

 $x'$ 

는

 $Q(\hat{x})$ 

일 수 있다.

[271] 따라서, 실시예의 아키텍처는, 2 개의 서브-네트워크들

 $I$ 

및

 $Q$ 

를 수학적 9의 총 손실

 $L$ 

을 최소화하도록 조인트되어 최적화될 수 있게 할 수 있다. 여기에서,

 $\hat{x}$ 

는

 $Q$ 

가 최종적인 재구축을 고 충실도(high fidelity)로 출력한다는 뜻에서 최적으로 표현될 수 있다.

[272] 실시예는 커스터마이징된 품질 향상 네트워크보다는, 이미지 압축 및 품질 향상의 양자에 대한 조인트 엔드-투-엔드 학습 스킴을 제시할 수 있다. 따라서, 적합한 품질 향상 네트워크를 선택하기 위해, 참조 이미지 압축 방법이 다양한 품질 향상 방법들과 캐시케이스 연결들로 결합될 수 있다.

[273] 실시예에서, 이미지 압축 네트워크는 품질 향상 네트워크들의 검증된(verified) 지혜들(wisdoms)을 활용할 수 있다. 품질 향상 네트워크의 검증된 지혜들은

슈퍼-레졸루션(super-resolution) 및 아티팩트-감축(artifact-reduction)을 포함할 수 있다. 예를 들면, 품질 향상 네트워크는 매우 깊은 슈퍼 레졸루션(Very Deep Super Resolution; VDSR), 잔차 밀도 네트워크(Residual Dense Network; RDN) 및 그룹된 잔차 밀도 네트워크(Grouped Residual Dense Network; GRDN)을 포함할 수 있다.

- [274] 도 9는 일 실시예에 따른 이미지 압축 네트워크의 전반적인 네트워크 아키텍처를 나타낸다.
- [275] 도 9는 자동 부호기(auto encoder)인 이미지 압축 네트워크의 아키텍처를 나타낼 수 있다. 자동 부호기의 구조는 부호기(encoder) 및 복호기(decoder)에 대응할 수 있다.
- [276] 말하자면, 부호기 및 복호기를 위해, 컨볼루션(convolution) 자동 부호기 구조가 사용될 수 있고, 분포 추정자

$f$

또한 컨볼루션 신경 네트워크들과 함께 구현될 수 있다.

- [277] 도 9 및 이하의 도면들에서는, 이미지 압축 네트워크의 아키텍처에 대하여, 아래와 같이 약어들 및 기호법(notation)들이 사용될 수 있다:

[278]

-

$g_a$   
는

$x$   
를 은닉 표현성분

$y$

로 변환하기 위한 분석 변환을 나타낼 수 있다.

[279]

-

$g_s$   
는 재구성된 이미지

$\hat{x}$

를 생성하기 위한 합성(synthesis) 변환을 나타낼 수 있다.

[280]

-

$h_a$   
는

$\hat{y}$   
의 공간적 중복성들을 은닉 표현성분

$z$

로 포착(capture)하기 위한 분석 변환을 나타낼 수 있다.

[281]

-

$h_s$   
는 모델 추정에 대한 문맥들을 생성하기 위한 합성 변환을 나타낼 수 있다.

- [282] - "conv"가 표시된 사각형은 컨볼루션 레이어를 나타낼 수 있다.
- [283] - 컨볼루션 레이어는 "필터들의 개수"  
 ×  
 "필터 높이"  
 ×  
 "필터 폭" / "다운-스케일링 또는 업-스케일링의 팩터(factor)"로서 표현될 수 있다.
- [284] - "  
 ↑  
 " 및 "  
 ↓  
 "는 트랜스포스된(transposed) 컨볼루션들을 통한 업-스케일링 및 다운-스케일링을 각각 나타낼 수 있다.
- [285] - 입력 이미지는 -1 및 1 사이의 스케일로 정규화될 수 있다.
- [286] - 컨볼루션 레이어에서 "N" 및 "M"은 특징 맵 채널(feature map channel)들의 개수들을 가리킬 수 있다. 반면, 각 완전-연결된(fully-connected) 레이어 내의 "M"은 노드들의 개수 및 부수하는(accompanying) 정수의 곱일 수 있다.
- [287] - "GDN"은 일반화된 분할 정규화(Generalized Divisive Normalization; GDN)를 나타낼 수 있다. "IGDN"은 역 일반화된 분할 정규화(Inverse Generalized Divisive Normalization; IGDN)를 나타낼 수 있다.
- [288] - "ReLU"는 렐루(relu) 레이어를 나타낼 수 있다.
- [289] - "Q"는 균일 양자화(반올림)을 나타낼 수 있다.
- [290] - "EC"는 엔트로피 부호화 프로세스을 나타낼 수 있다. "ED"는 엔트로피 복호화 프로세스을 나타낼 수 있다.
- [291] - "정규화(normalization)"는 정규화를 나타낼 수 있다.
- [292] - "비정규화(denormalization)"는 비정규화를 나타낼 수 있다.
- [293] - "abs"는 절대(absolute) 연산자(operator)을 나타낼 수 있다.
- [294] - "exp"는 자승(exponentiation) 연산자를 나타낼 수 있다.
- [295] - "  
 $f$   
 "는 모델 파라미터 추정자를 나타낼 수 있다.
- [296] - "  
 $E'$   
 ,  
 $E''$   
 및  
 $E'''$   
 는 3 개의 타입들의 문맥들을 추출하기 위한 함수(function)들을 각각 나타낼 수

있다.

- [297] 이미지 압축 네트워크에서, 콘볼루션 신경 네트워크들은 변환 및 재구축 기능들을 구현하기 위해 사용될 수 있다.
- [298] 도 9에서 도시된 것과 같이, 이미지 압축 네트워크 및 품질 향상 네트워크는 캐시케이드로 연결될 수 있다. 예를 들면, 품질 향상 네트워크는 GRDN일 수 있다.
- [299] 율-왜곡(rate-distortion) 최적화 및 변환 함수들에 관해 전술된 설명들이 실시예에 적용될 수 있다.
- [300] 이미지 압축 네트워크는 입력 이미지  $x$  를 은닉 표현성분들  $y$  로 변환할 수 있다. 다음으로,  $y$  는  $\hat{y}$  로 양자화될 수 있다.
- [301] 이미지 압축 네트워크는 하이퍼프라이어(hyperprior)  $\hat{z}$  를 사용할 수 있다.  $\hat{z}$  는  $\hat{y}$  의 공간적(spatial) 상관관계들(correlations)을 포착(capture)할 수 있다.
- [302] 이미지 압축 네트워크는 4 개의 근본적인 변환 함수들을 사용할 수 있다. 변환 함수들은 전술된 분석 변환  $g_a(\mathbf{x}; \phi_g)$  , 합성 변환  $g_s(\hat{\mathbf{y}}; \theta_g)$  , 분석 변환  $h_a(\hat{\mathbf{y}}; \phi_h)$  및 합성 변환  $h_s(\hat{\mathbf{z}}; \theta_h)$  일 수 있다.
- [303] 도 9에서 도시된  $g_a$  ,

$g_s$

,

$h_a$

및

$h_s$

에 대해서 전술된 다른 실시예에서의 설명이 적용될 수 있다. 또한,

$h_s$

의 말단(end)에서는, 절대(absolute) 연산자(operator)가 아닌 자승(exponentiation) 연산자가 사용될 수 있다.

[304] 실시예의 율-왜곡에 대한 최적화 프로세스는 이미지 압축 네트워크가

$\hat{y}$

및

$\hat{z}$

의 엔트로피를 가능한 낮게 도출(yield)하는 것을 보장할 수 있다. 또한, 최적화 프로세스는 이미지 압축 네트워크가

$\hat{y}$

로부터 재구축되는 출력 이미지

$x'$

를 가능한 원래의 시각적(visual) 품질에 근접하도록 도출하는 것을 보장할 수 있다.

[305] 이러한 율-왜곡 최적화를 위해, 입력 이미지

$x$

및 출력 이미지

$x'$

간의 왜곡이 계산될 수 있다, 율(rate)은

$\hat{y}$

및

$\hat{z}$

에 대한 사전 확률 모델들(prior probability models)에 기반하여 계산될 수 있다.

[306]  $\hat{z}$

에 대하여,

$\mathcal{U}(-\frac{1}{2}, \frac{1}{2})$

와 콘볼브된 단순(simple) 제로-평균(zero-mean) 가우시안 모델이 사용될 수 있다. 단순 제로-평균 가우시안 모델의 표준 편차들은 훈련을 통해 갖춰질 수 있다. 반면, 전술된 실시예에서 설명된 것과 같이,

$\hat{y}$

에 대한 사전 확률 모델은 모델 파라미터 추정자

$f$

- 에 의해 자동회귀 방식(auto-regressive manner)으로 추정될 수 있다.
- [307] 전술된 실시예에서 설명된 것과 같이, 모델 파라미터 추정자  $f$  는 2 개의 타입들의 문맥들을 활용할 수 있다.
- [308] 2 개의 타입들의 문맥들은 비트-소비(bit-consuming) 문맥  $c'_i$  및 비트-프리(bit-free) 문맥  $c''_i$  일 수 있다.  $c'_i$  는 하이퍼프라이어  $\hat{z}$  로부터 재구축될 수 있다.  $c''_i$  는  $\hat{y}$  의 인접한 알려진 표현성분들로부터 추출될 수 있다.
- [309] 추가하여, 실시예에서, 모델 파라미터 추정자  $f$  는 모델 파라미터들을 더 정교하게 추정하기 위해 전역 문맥  $c'''_i$  를 활용할 수 있다.
- [310] 3 개의 주어진 문맥들을 가지고,  $f$  는 (  $\mathcal{U}(-\frac{1}{2}, \frac{1}{2})$  와 콘볼브된) 가우시안 혼합 모델(Gaussian Mixture Model; GMM)의 파라미터들을 추정할 수 있다. 실시예에서, GMM은  $\hat{y}$  에 대한 사전 확률 모델로서 채용될 수 있다. 이러한 파라미터 추정은 EC 및 ED로 표현된 엔트로피 부호화 프로세스 및 엔트로피 복호화 프로세스에서 사용될 수 있다. 또한, 파라미터 추정은 훈련을 위한 율 항(rate term)의 계산에서도 사용될 수 있다.
- [311] 도 10은 일 예에 따른 모델 파라미터 추정자의 구조를 나타낼 수 있다.
- [312] 도 11은 일 예에 따른 비-로컬 문맥 프로세싱 네트워크를 나타낼 수 있다.
- [313] 도 12는 일 예에 따른 오프셋-문맥 프로세싱 네트워크를 나타낼 수 있다.
- [314] 도 10, 도 11 및 도 12에서는, 이미지 압축 네트워크의 아키텍처에 대하여,

아래와 같이 약어들 및 기보법(notation)들이 사용될 수 있다:

- [315] - "FCN"은 완전-연결된 네트워크(fully-connected network)를 나타낼 수 있다.  
 [316] - "concat"는 연쇄(concatenation) 연산자를 나타낼 수 있다.  
 [317] - "leakyReLU"는 유출되는(leaky) ReLU를 나타낼 수 있다. 유출되는 ReLU는 ReLU의 변형인 함수일 수 있으며, 유출되는(leaky) 정도가 특정되는 함수일 수 있다. 예를 들면, leakyReLU 함수에 대해 제1 설정 값 및 제2 설정 값이 설정될 수 있다. leakyReLU 함수는 입력 값이 제1 설정 값의 이하인 경우, 제1 설정 값을 출력하지 않고, 입력 값 및 제2 설정 값을 출력할 수 있다.

- [318] 모델 파라미터 추정자

$f$

의 구조는

$f$

를 새로운 모델 추정자로 확장함으로써 향상될 수 있다. 새로운 모델 추정자는 모델 파라미터 추정의 능력(capability)을 향상시키기 위해 모델 파라미터 개선 모듈(Model Parameter Refinement Module; MPRM)을 접목할 수 있다.

- [319] MPRM은 2 개의 잔차(residual) 블록들을 가질 수 있다. 2 개의 잔차(residual) 블록은 오프셋-문맥(offset-context) 프로세싱 네트워크 및 비-로컬(non-local) 문맥 프로세싱 네트워크일 수 있다.

- [320] 2 개의 잔차 블록들의 각각은 완전-연결된(fully-connected) 레이어들 및 대응하는(corresponding) 비-선형(non-linear) 활성화(activation) 레이어들을 포함할 수 있다.

- [321] 엔트로피-최소화에 대한 향상된 엔트로피 모델 및 파라미터 추정

- [322] 전술된 실시예의 엔트로피-최소화 방법은 각

$\hat{y}_i$

에 대한 사전 모델 파라미터들을 추정하기 위해 로컬 문맥들을 활용할 수 있다. 엔트로피-최소화 방법은 현재의 은닉 표현성분

$\hat{y}_i$

에 대한 (균일 함수(uniform function)과 콘볼드된) 단일(single) 가우시안 사전 모델(Gaussian prior model)의 표준 편차 파라미터

$\sigma_i$

및 평균 파라미터

$\mu_i$

를 추정하기 위해 현재의 은닉 표현성분

$\hat{y}_i$

의 이웃 은닉 표현성분들을 활용할 수 있다.

- [323] 이러한 접근방식들은 아래의 2 개의 제한들을 가질 수 있다.

- [324] (i) 단일 가우시안 모델은 은닉 표현성분들의 다양한 분포를 모델링하는데 있어서 제한된 능력을 가질 수 있다. 실시예에서는, 가우시안 혼합 모델(Gaussian

Mixture Model; GMM)이 사용될 수 있다.

[325] (ii) 이웃 은닉 표현성분들의 상관관계(correlation)들이 전체의 공간의 도메인들(spatial domains)에 걸쳐 널리 퍼져 존재할 때에는 이웃 은닉 표현성분들로부터 문맥 정보를 추출하는 것이 제한될 수 있다.

[326] 사전 분배들에 대한 가우시안 혼합 모델

[327] 전술된 실시예의 자동회귀 접근방식들은 각

$\hat{y}_i$

의 분포를 모델링하기 위해 단일 가우시안 분포(또는, 가우시안 사전 모델)를 사용할 수 있다. 이러한 자동회귀 방법들의 변환 네트워크들이 단일 가우시안 분포들을 따르는 은닉 표현성분들을 생성할 수 있지만, 이러한 단일 가우시안 모델링은 은닉 표현성분들의 실제의 분포들을 예측함에 있어서 제한될 수 있으며, 따라서 차선의(sub-optimal) 성능으로 이끌 수 있다. 대신, 실시예에서는 더 일반화된 형태인 사전 확률 모델의 GMM이 사용될 수 있다. GMM은 실제의 분포들을 더 정확하게 근사할 수 있다.

[328] 아래의 수학적 식 10은 GMM을 사용하는 엔트로피 모델을 나타낼 수 있다.

[329] [수식10]

$$p_{\tilde{y}|\hat{z}}(\tilde{y} | \hat{z}, \theta_h) = \prod_i \left( \sum_{g=1}^G \phi_{i,g} \mathcal{N}(\mu_{i,g}, \sigma_{i,g}^2) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right) (\tilde{y}_i)$$

$$\text{with } \{\mu_{i,g}, \sigma_{i,g} | 1 \leq g \leq G\} = f(c'_i, c''_i),$$

$$c'_i = E'(h_s(\hat{z}; \theta_h), i),$$

$$c''_i = \{E''(\langle \hat{y} \rangle, i), E'''(\langle \hat{y} \rangle, i), o_i\}$$

$$\hat{z} = Q(z)$$

[330] 엔트로피 모델들에 대한 공식(formulation)

[331] 기본적으로, 전술된 실시예의 수학적 식 9를 참조하여 전술된 R-D 최적화 프레임워크가 실시예의 엔트로피 모델을 위해 사용될 수 있다.

[332] 윗 항은

$\tilde{z}$

및

$\tilde{y}|\hat{z}$

에 대한 크로스-엔트로피로 구성될 수 있다.

[333] 양자화에 인한 불연속성(discontinuity)들을 다루기 위해서,

$\hat{y}$

의 확률 질량 함수(Probability Mass Function; PMF)를 근사하기 위해 균일 함수  $\mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$

와 콘볼브된 밀도 함수가 사용될 수 있다. 따라서, 훈련에 있어서, 노이즈 낀 표현성분들

$\tilde{y}$

및

$\tilde{z}$

가 실제의 샘플 분포들을 PMF-근사 함수들로 피트(fit)시키기 위해 사용될 수 있다. 여기에서,

$\tilde{y}$

및

$\tilde{z}$

는 균일 분포들을 따를 수 있고,

$\tilde{y}$

의 평균 값은

$y$

일 수 있고,

$\tilde{z}$

의 평균 값은

$z$

일 수 있다.

[334]  $\tilde{z}$

의 분포를 모델링하기 위해, 전술된 실시예에서 설명된 것과 같이, (균일 밀도 함수와 콘볼브된) 제로-평균(zero-mean) 가우시안 밀도 함수들이 사용될 수 있다. 제로-평균 가우시안 밀도 함수들의 표준 편차들은 훈련을 통해 최적화될 수 있다.

[335]  $\tilde{y}|\tilde{z}$

에 대한 엔트로피 모델은 GMM에 기반하여 아래의 수학식 11 및 수학식 13와 같이 확장될 수 있다.

[336] [수식11]

$$p_{\tilde{y}|\tilde{z}}(\tilde{y} | \tilde{z}, \theta_h) = \prod_i \left( \sum_{g=1}^G \phi_{i,g} \mathcal{N}(\mu_{i,g}, \sigma_{i,g}^2) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right) (\tilde{y}_i)$$

with  $\{\mu_{i,g}, \sigma_{i,g} | 1 \leq g \leq G\} = f(c'_i, c''_i)$ ,

$c'_i = E'(h_s(\tilde{z}; \theta_h), i)$ ,

$c''_i = \{E''(\langle \hat{y} \rangle, i), E'''(\langle \hat{y} \rangle, i), o_i\}$

$\tilde{z} = Q(z)$

[337] 수학식 11에서, 아래의 수학식 12는 가우시안 혼합을 나타낼 수 있다.

[338] [수식12]

$$\sum_{g=1}^G \phi_{i,g} \mathcal{N}(\mu_{i,g}, \sigma_{i,g}^2)$$

[339] 수학식 11에서,

$$E'''(\langle \hat{y} \rangle, i)$$

는 비-로컬 문맥들을 나타낼 수 있다.

[340] 수학식 11에서,

$$o_i$$

는 오프셋들을 나타낼 수 있다. 오프셋은 원-핫 코드될(one-hot coded) 수 있다.

[341] 수학식 11은 병합된 모델의 공식을 나타낼 수 있다. 구조적인 변경들은 수학식 11에 따른 모델 공식과는 무관할 수 있다.

[342] [수식13]

$$p_{\tilde{y}|\hat{z}}(\tilde{y}|\hat{z}) = \prod_i \left( \sum_{g=1}^G \pi_{i,g} \mathcal{N}(\mu_{i,g}, \sigma_{i,g}^2) * \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right) \right) (\tilde{y}_i)$$

with  $\{\pi_{i,g}, \mu_{i,g}, \sigma_{i,g} | 1 \leq g \leq G\} = f(\mathbf{c}'_i, \mathbf{c}''_i, \mathbf{c}'''_i)$

[343]  $G$

는 가우시안 분포 함수들의 개수일 수 있다.

[344] 모델 파라미터 추정자

$$f$$

는

$$G$$

개의 파라미터들을 예측할 수 있고, 예측을 통해

$$G$$

개의 가우시안 분포들의 각 가우시안 분포가 그 자신의 가중치(weight)

파라미터

$$\pi_{i,g}$$

, 평균 파라미터

$$\mu_{i,g}$$

및 표준 편차 파라미터

$$\sigma_{i,g}$$

를 가지게 할 수 있다.

[345] 평균 제곱된 에러(Mean Squared Error; MSE)는 전술된 수학식 9에서의 최적화를 위하여 왜곡 향으로서 기본적으로 사용될 수 있다. 또한, 왜곡 향으로서 다중스케일 구조적 유사도(MultiScale-Structural SIMilarity; MS-SSIM) 최적화된 모델이 사용될 수 있다.

- [346] 모델 파라미터 추정을 위한 전역 문맥
- [347] 도 13은 일 예에 따른 전역 문맥 지역에 매핑된 변수들을 나타낸다.
- [348] 현재의 은닉 표현성분에 대한 문맥 정보를 더 잘 추출하기 위해, 사전 모델 파라미터들을 추정하기 위한 알려진 표현성분들의 전체의 영역(area)으로부터 모든 가능한 문맥들을 집계(aggregating)함으로써 전역 문맥이 사용될 수 있다.
- [349] 전역 문맥의 사용을 위해, 전역 문맥은 로컬 문맥 지역(region) 및 비-로컬 문맥 지역으로부터 집계된 정보로서 정의될 수 있다.
- [350] 이하에서, 용어들 "영역(area)" 및 "지역(region)"은 동일한 의미로 사용될 수 있고, 서로 교체되어 사용될 수 있다.
- [351] 여기에서, 로컬 문맥 지역은 현재의 은닉 표현성분  $y_i$ 로부터 고정된 거리 내의 지역일 수 있다.  
 $K$   
 는 고정된 거리를 나타낼 수 있다. 비-로컬 문맥 지역은 로컬 문맥 지역의 외부의(outside) 전체의 인과관계의(causal) 영역일 수 있다.
- [352] 전역 문맥  $c_i'''$   
 로서, 전역 문맥 지역으로부터 집계된 가중치가 부여된(weighted) 평균 값 및 가중치가 부여된 표준 편차 값이 사용될 수 있다.
- [353] 전역 문맥 지역은  $\dot{y}$   
 의 채널 내의 전체의 알려진 공간적 영역일 수 있다.  
 $\dot{y}$   
 는 1  
 $\times$   
 1 콘볼루션 레이어를 통한  $\hat{y}$   
 의 선형으로(linearly) 변환된 버전일 수 있다.
- [354] 전역 문맥  $c_i'''$   
 은,  
 $\hat{y}$   
 로부터 보다는,  
 $\hat{y}$   
 의 서로 다른 채널들에 걸친 상관관계들을 또한 포착하기 위해  $\dot{y}$   
 로부터 획득될 수 있다.

[355] 전역 문맥

$$c_i'''$$

은 아래의 수학식 14와 같이 표현될 수 있다.

[356] [수식14]

$$c_i''' = \{\mu_i^*, \sigma_i^*\}$$

[357] 전역 문맥

$$c_i'''$$

은 가중치가 부여된 평균

$$\mu_i^*$$

및 가중치가 부여된 표준 편차

$$\sigma_i^*$$

를 포함할 수 있다.

[358]  $\mu_i^*$

는 아래의 수학식 15와 같이 정의될 수 있다.

[359] [수식15]

$$\mu_i^* = \sum_{k,l \in S} w_{k,l}^{(i)} \dot{y}_{i_h-k, i_v-l}^{(i)}$$

[360]  $\sigma_i^*$

는 아래의 수학식 16과 같이 정의될 수 있다.

[361] [수식16]

$$\sigma_i^* = \sqrt{\frac{\sum_{k,l \in S} w_{k,l}^{(i)} (\dot{y}_{i_h-k, i_v-l}^{(i)} - \mu_i^*)^2}{1 - \sum_{k,l \in S} w_{k,l}^{(i)2}}}$$

[362]  $i$

는 아래의 수학식 17과 같이 정의될 수 있다.

[363] [수식17]

$$i = [i_c, i_h, i_v]$$

[364]  $i$

는

$$i_c$$

번째 채널 내에서 현재의 위치

$$(i_h, i_v)$$

를 가리키는 3-차원 시공간-채널-별(spatio-channel-wise) 위치(position)

인덱스일 수 있다.

[365]  $w_{k,l}^{(i)}$

는 현재의 위치

$(i_h, i_v)$

에 기반한 상대적 좌표들

$(k, l)$

에 대한 가중치 변수일 수 있다.

[366]  $\dot{y}_{i_h-k, i_v-l}^{(i)}$

는 전역 문맥 지역

$S$

내에서, 위치

$(i_h-k, i_v-l)$

에서의

$\dot{y}^{(i)}$

의 표현성분일 수 있다.

[367]  $\dot{y}^{(i)}$

는

$\dot{y}$

의

$i_c$

번째 채널 내에서의 2-차원 표현성분들일 수 있다.

[368]  $w^{(i)}$

내의 가중치 변수들은 정규화된 가중치들일 수 있다. 정규화된 가중치들은 요소-별로(element-wise)

$\dot{y}^{(i)}$

에 곱해질 수 있다. 수학식 15에서, 가중치 변수들은 가중치가 부여된 평균을 위하여 요소 별로

$\dot{y}^{(i)}$

에 곱해질 수 있다. 수학식 16에서, 가중치 변수들은

$(\dot{y}_{i_h-k, i_v-l}^{(i)} - \mu_i^*)$

의 차이 제곱(difference square)들로 곱해질 수 있다.

[369] 실시예에서, 키 이슈는 모든 위치

$i$

에서 가중치 변수들

$w^{(i)}$

의 최적의 세트를 발견하는 것일 수 있다. 고정된 개수의 훈련가능한 변수들

$\psi^{(i)}$

로부터

$w^{(i)}$

를 획득하기 위해,

$w^{(i)}$

는 2-차원 확장(extension)에서 1-차원 전역 문맥 지역을 추출하는 스킴에  
기반하여 추정될 수 있다.

[370] 도 13에서는, 1) 고정된 거리

$K$

내의 로컬 문맥 지역 및 2) 가변의 크기를 갖는 비-로컬 문맥 지역을 포함하는  
전역 문맥 지역이 도시된다.

[371] 로컬 문맥 지역은 훈련가능한 변수들

$\psi^{(i)}$

에 의해 커버될 수 있다. 비-로컬 문맥 지역은 로컬 문맥 영역의 외부(outside)일  
수 있다.

[372] 전역 문맥 추출에서, 비-로컬 문맥 지역은 로컬 문맥 영역을 정의하는 지역  
문맥 윈도우가 특징 맵 상으로 슬라이드됨에 따라 확대될 수 있다. 비-로컬 문맥  
지역이 확대됨에 따라 가중치 변수들

$w^{(i)}$

의 개수는 증가할 수 있다.

[373] 훈련가능한 변수들

$\psi^{(i)}$

의 고정된 크기에 의해 커버될 수 없는 비-로컬 문맥 지역을 다루기 위해, 도  
13에서 도시된 것과 같이, 가장 가까운 로컬 문맥 지역에 할당된

$\psi^{(i)}$

의 변수가 비-로컬 문맥 지역 내의 각 공간적 위치에 대해 사용될 수 있다.

[374] 그 결과로서, 훈련가능한 변수들

$\psi^{(i)}$

의 집합

$a^{(i)}$

이 획득될 수 있다.

$a^{(i)}$

는 전역 문맥 지역에 대응할 수 있다.

[375] 다음으로,

$w^{(i)}$

는 아래의 수학적 식 18 같이 소프트맥스(softmax)를 통해

$a^{(i)}$

를 정규화함으로써 계산될 수 있다.

[376] [수식18]

$$\mathbf{w}^{(i)} = \text{softmax}(\mathbf{a}^{(i)})$$

[377]  $\mathbf{a}^{(i)}$

는 아래의 수학적식 19과 같이 정의될 수 있다.

[378] [수식19]

$$\mathbf{a}^{(i)} = \{\psi_{clip(k,K),clip(l,K)}^{(i)} | k, l \in S\}$$

[379]  $clip(x, K)$

는 아래의 수학적식 20과 같이 정의될 수 있다.

[380] [수식20]

$$clip(x, K) = \max(-K, \min(K, x))$$

[381] 동일한 채널 내에서(즉, 동일한 공간적 특징 공간에 걸쳐) 아래의 수학적식 21이 성립할 수 있다.

[382] [수식21]

$$\psi_{k,l}^{(i)} = \psi_{k,l}^{(i+c)}$$

[383]  $\dot{\psi}$

의 몇 개의 채널들에 대하여, 훈련된

$$\psi^{(i)}$$

의 예들이 시각화될 수 있다. 예를 들면, 채널의 문맥은 현재 은닉 표현성분의 바로 옆에 있는 이웃 표현성분에 의존할 수 있다. 또는, 채널의 문맥은 넓게 확산된 이웃 표현성분들에 의존할 수 있다.

[384] 도 14는 일 실시예에 따른 GRDN의 구조를 나타낸다.

[385] 구현에서는, 중간(intermediate) 재구축이 GRDN으로 입력될 수 있고, 최종의 재구축이 GRDN으로부터 출력될 수 있다.

[386] 도 14에서는, GRDN의 아키텍처에 대하여, 아래와 같이 약어들 및 기호법(notation)들이 사용될 수 있다:

[387] - "GRDB"는 그룹된 잔차 밀도 블록(Grouped Residual Dense Block; GRDB)를 나타낼 수 있다.

[388] - "CBAM"은 컨볼루션 블록 어텐션 모듈(Convolutional Block Attention Module; CBAM)을 나타낼 수 있다.

[389] - "Conv. Up"은 컨볼루션 업-샘플링을 나타낼 수 있다.

[390] - "+"는 합(addition) 연산을 나타낼 수 있다.

[391] 도 15는 일 실시예에 따른 GRDN의 GRDB의 구조를 나타낸다.

[392] 도 15에서는, GRDB의 아키텍처에 대하여, 아래와 같이 약어들 및 기호법(notation)들이 사용될 수 있다:

[393] - "RDB"는 잔차 밀도 블록(Residual Dense Block; RDB)을 나타낼 수 있다.

[394] 도 16은 일 실시예에 따른 GRDB의 RDB의 구조를 나타낸다.

- [395] 도 14, 도 15 및 도 16을 참조하여 예시된 것과 같이, GRDN의 구현을 위해 4개의 GRDB들이 사용될 수 있다. 또한, 각 GRDB를 위해 3개의 RDB들이 사용될 수 있다. 각 RDB를 위해 3개의 콘볼루션 레이어들이 사용될 수 있다.
- [396] 부호기-복호기 모델
- [397] 도 17은 실시예에 따른 부호기를 나타낸다.
- [398] 도 17에서, 우측의 작은 아이콘들은 엔트로피-부호화된 비트스트림을 나타낼 수 있다.
- [399] 도 17에서, EC는 엔트로피 코딩(즉, 엔트로피 인코딩)을 나타낼 수 있다.
- U | Q
- 는 균일 잡음 추가 또는 균일 양자화를 나타낼 수 있다.
- [400] 또한, 도 17에서, 노이즈가 낀 표현성분들은 점선들(dotted lines)로 도시되었다. 실시예에서, 노이즈가 낀 표현성분들은 엔트로피 모델들로의 입력으로서 단지 훈련을 위해 사용될 수 있다.
- [401] 도 17에서 도시된 것과 같이, 부호기는 도 9를 참조하여 전술된 자동 부호기에서 부호화 프로세스에 대한 요소들을 포함할 수 있고, 자동 부호기의 부호화를 수행할 수 있다. 말하자면, 실시예의 부호기는 도 9를 참조하여 전술된 자동 부호기가 입력 이미지에 대한 부호화를 수행하는 측면으로 보인 것일 수 있다.
- [402] 따라서, 도 9를 참조하여 전술된 자동 부호기에 대한 설명은 본 실시예의 부호기에도 적용될 수 있다.
- [403] 부호기 및 복호기의 동작들 및 상호작용(interaction)에 대해서 아래에서 더 상세하게 설명된다.
- [404] 도 18은 실시예에 따른 복호기를 나타낸다.
- [405] 도 18에서, 좌측의 작은 아이콘들은 엔트로피-부호화된 비트스트림을 나타낼 수 있다.
- [406] ED는 엔트로피 디코딩을 나타낼 수 있다.
- [407] 도 18에서 도시된 것과 같이, 복호기는 도 9를 참조하여 전술된 자동 부호기에서 복호화 프로세스에 대한 요소들을 포함할 수 있고, 자동 부호기의 복호화를 수행할 수 있다. 말하자면, 실시예의 복호기는 도 9를 참조하여 전술된 자동 복호기가 입력 이미지에 대한 복호화를 수행하는 측면으로 보인 것일 수 있다.
- [408] 따라서, 도 9를 참조하여 전술된 자동 부호기에 대한 설명은 본 실시예의 복복호기에도 적용될 수 있다.
- [409] 부호기 및 복호기의 동작들 및 상호작용에 대해서 아래에서 더 상세하게 설명된다.
- [410] 부호기는 입력 이미지를 은닉 표현성분들로 변환할 수 있다. 부호기는 은닉 표현성분들을 양자화함으로써 양자화된 은닉 표현성분들을 생성할 수 있다. 또한, 부호기는 양자화된 은닉 표현성분들에 대해 훈련된 엔트로피 모델을

사용하는 엔트로피-부호화를 수행함으로써 엔트로피-인코딩된 은닉 표현성분들을 생성할 수 있고, 엔트로피-부호화된 은닉 표현성분들을 비트스트림으로서 출력할 수 있다.

[411] 훈련된 엔트로피 모델은 부호기 및 복호기 간에 공유될 수 있다. 말하자면, 훈련된 엔트로피 모델은 공유된 엔트로피 모델로도 칭해질 수 있다.

[412] 반면, 복호기는 비트스트림을 통해 엔트로피-부호화된 은닉 표현성분들을 수신할 수 있다. 복호기는 엔트로피-인코딩된 은닉 표현성분들에 대해 공유된 엔트로피 모델을 사용하는 엔트로피-디코딩을 수행함으로써 은닉 표현성분들을 생성할 수 있다. 복호기는 은닉 표현성분들을 사용하여 재구성된 이미지를 생성할 수 있다.

[413] 부호기 및 복호기에 있어서, 모든 파라미터들은 이미 훈련된 것으로 가정될 수 있다.

[414] 부호기-복호기 모델의 구조(structure)는 기본적으로

$g_a$

및

$g_s$

를 포함할 수 있다.

$g_a$

는

$x$

의

$y$

로의 변환을 담당할 수 있으며,

$g_s$

는

$g_a$

의 변환에 대한 역변환(inverse transform)을 담당할 수 있다.

[415] 변환된

$y$

는 라운딩에 의해

$\hat{y}$

로 균일하게 양자화될 수 있다.

[416] 여기에서, 기존의 코덱들과는 다르게, 엔트로피 모델들에 기반한 접근방식들의 경우에는, 표현성분들의 스케일들이 훈련에 의해 함께 최적화되기 때문에 양자화 스텝들에 대한 튜닝은 일반적으로 불필요할 수 있다.

[417]  $g_a$

및

$g_s$

의 사이의 다른 구성요소들은 1) 공유된 엔트로피 모델들 및 2) 기저에 있는(underlying) 문맥 준비(preparation) 프로세스들을 가지고 엔트로피 부호화(또는, 엔트로피 복호화)의 역할을 수행할 수 있다.

[418] 보다 구체적으로, 엔트로피 모델은 각

 $\hat{y}_i$ 

의 분포를 개별적으로 추정할 수 있다. 각

 $\hat{y}_i$ 

의 분포의 추정에 있어서,

 $\pi_i$ 

,

 $\mu_i$ 

및

 $\sigma_i$ 

는 주어진 문맥들의 3 개의 타입들인

 $c'_i$ 

,

 $c''_i$ 

및

 $c'''_i$ 

을 가지고 추정될 수 있다.

[419] 이러한 문맥들 중에서,

 $c'$ 

는 추가의 비트 할당을 요구하는 부가 정보일 수 있다.

 $c'$ 

를 운반하기 위해 요구되는 비트-레이트를 감소시키기 위해,

 $\hat{y}$ 

로부터 변환된 은닉 표현성분

 $z$ 

는

 $z$ 

자신의 엔트로피 모델에 의해 양자화 및 엔트로피-부호화될 수 있다.

[420] 반면,

 $c''_i$ 

는 어떤 추가의 비트 할당 없이

 $\langle \hat{y} \rangle$ 

로부터 추출될 수 있다. 여기에서,

 $\langle \hat{y} \rangle$

는 엔트로피 부호화 또는 엔트로피 복호화 진행함에 따라 변할 수 있다. 그러나,  
 $\langle \hat{y} \rangle$

는 동일한

$\hat{y}_i$

를 처리함에 있어서 언제나 부호기 및 복호기의 양자 내에서 동일할 수 있다.

[421]  $c_i'''$

는

$\dot{y}$

로부터 추출될 수 있다.

$h_s$

의 파라미터들 및 엔트로피 모델들은 부호기 및 복호기의 양자에 의해  
 단순하게 공유될 수 있다.

[422] 훈련이 진행되는 동안 엔트로피 모델들로의 입력들은 노이즈 낀 표현성분들일 수 있다. 노이즈 낀 표현성분들은 엔트로피 모델이 이산 표현성분들의 확률 질량 함수들에 근사하도록 할 수 있다.

[423] 도 19는 일 실시예에 따른 부호화 장치의 구조도이다.

[424] 부호화 장치(1900)는 버스(1990)를 통하여 서로 통신하는 처리부(1910), 메모리(1930), 사용자 인터페이스(User Interface; UI) 입력 디바이스(1950), UI 출력 디바이스(1960) 및 저장소(storage)(1940)를 포함할 수 있다. 또한, 부호화 장치(1900)는 네트워크(1999)에 연결되는 통신부(1920)를 더 포함할 수 있다.

[425] 처리부(1910)는 중앙 처리 장치(Central Processing Unit; CPU), 메모리(1930) 또는 저장소(1940)에 저장된 프로세싱(processing) 명령어(instruction)들을 실행하는 반도체 장치일 수 있다. 처리부(1910)는 적어도 하나의 하드웨어 프로세서일 수 있다.

[426] 처리부(1910)는 장치(1900)로 입력되거나, 장치(1900)에서 출력되거나, 장치(1900)의 내부에서 사용되는 신호, 데이터 또는 정보의 생성 및 처리를 수행할 수 있고, 신호, 데이터 또는 정보에 관련된 검사, 비교 및 판단 등을 수행할 수 있다. 말하자면, 실시예에서 데이터 또는 정보의 생성 및 처리와, 데이터 또는 정보에 관련된 검사, 비교 및 판단은 처리부(1910)에 의해 수행될 수 있다.

[427] 처리부(1910)를 구성하는 요소들의 적어도 일부는 프로그램 모듈들일 수 있으며, 외부의 장치 또는 시스템과 통신할 수 있다. 프로그램 모듈들은 운영 체제, 응용 프로그램 모듈 및 기타 프로그램 모듈의 형태로 부호화 장치(1900)에 포함될 수 있다.

[428] 프로그램 모듈들은 물리적으로는 여러 가지 공지의 기억 장치 상에 저장될 수 있다. 또한, 이러한 프로그램 모듈 중 적어도 일부는 부호화 장치(1900)와 통신 가능한 원격 기억 장치에 저장될 수도 있다.

[429] 프로그램 모듈들은 일 실시예에 따른 기능 또는 동작을 수행하거나, 일

실시예에 따른 추상 데이터 유형을 구현하는 루틴(routine), 서브루틴(subroutine), 프로그램, 오브젝트(object), 컴퍼넌트(component) 및 데이터 구조(data structure) 등을 포괄할 수 있지만, 이에 제한되지는 않는다.

- [430] 프로그램 모듈들은 부호화 장치(1900)의 적어도 하나의 프로세서(processor)에 의해 수행되는 명령어(instruction) 또는 코드(code)로 구성될 수 있다.
- [431] 처리부(1910)는 전술된 부호화기에 대응할 수 있다. 말하자면, 도 17을 참조하여 전술된 부호화기 및 도 9를 참조하여 전술된 자동 부호기의 부호화에 대한 동작은 처리부(1910)에 의해 수행될 수 있다.
- [432] 저장부는 메모리(1930) 및/또는 저장소(1940)를 나타낼 수 있다. 메모리(1930) 및 저장소(1940)는 다양한 형태의 휘발성 또는 비휘발성 저장 매체일 수 있다. 예를 들면, 메모리(1930)는 롬(ROM)(1931) 및 램(RAM)(1932) 중 적어도 하나를 포함할 수 있다.
- [433] 저장부는 부호화 장치(1900)의 동작을 위해 사용되는 데이터 또는 정보를 저장할 수 있다. 실시예에서, 부호화 장치(1900)가 갖는 데이터 또는 정보는 저장부 내에 저장될 수 있다.
- [434] 부호화 장치(1900)는 컴퓨터에 의해 독출(read)될 수 있는 기록 매체를 포함하는 컴퓨터 시스템에서 구현될 수 있다.
- [435] 기록 매체는 부호화 장치(1900)가 동작하기 위해 요구되는 적어도 하나의 모듈을 저장할 수 있다. 메모리(1930)는 적어도 하나의 모듈을 저장할 수 있고, 적어도 하나의 모듈이 처리부(1910)에 의하여 실행되도록 구성될 수 있다.
- [436] 부호화 장치(1900)의 데이터 또는 정보의 통신과 관련된 기능은 통신부(1920)를 통해 수행될 수 있다.
- [437] 네트워크(1999)는 부호화 장치(1900) 및 복호화 장치(2000) 간의 통신을 제공할 수 있다.
- [438] 도 20은 일 실시예에 따른 복호화 장치의 구조도이다.
- [439] 복호화 장치(2000)는 버스(2090)를 통하여 서로 통신하는 처리부(2010), 메모리(2030), 사용자 인터페이스(User Interface; UI) 입력 디바이스(2050), UI 출력 디바이스(2060) 및 저장소(storage)(2040)를 포함할 수 있다. 또한, 복호화 장치(2000)는 네트워크(2099)에 연결되는 통신부(2020)를 더 포함할 수 있다.
- [440] 처리부(2010)는 중앙 처리 장치(Central Processing Unit; CPU), 메모리(2030) 또는 저장소(2040)에 저장된 프로세싱(processing) 명령어(instruction)들을 실행하는 반도체 장치일 수 있다. 처리부(2010)는 적어도 하나의 하드웨어 프로세서일 수 있다.
- [441] 처리부(2010)는 장치(2000)로 입력되거나, 장치(2000)에서 출력되거나, 장치(2000)의 내부에서 사용되는 신호, 데이터 또는 정보의 생성 및 처리를 수행할 수 있고, 신호, 데이터 또는 정보에 관련된 검사, 비교 및 판단 등을 수행할 수 있다. 말하자면, 실시예에서 데이터 또는 정보의 생성 및 처리와, 데이터 또는 정보에 관련된 검사, 비교 및 판단은 처리부(2010)에 의해 수행될 수

있다.

- [442] 처리부(2010)를 구성하는 요소들의 적어도 일부는 프로그램 모듈들일 수 있으며, 외부의 장치 또는 시스템과 통신할 수 있다. 프로그램 모듈들은 운영 체제, 응용 프로그램 모듈 및 기타 프로그램 모듈의 형태로 복호화 장치(2000)에 포함될 수 있다.
- [443] 프로그램 모듈들은 물리적으로는 여러 가지 공지의 기억 장치 상에 저장될 수 있다. 또한, 이러한 프로그램 모듈 중 적어도 일부는 복호화 장치(2000)와 통신 가능한 원격 기억 장치에 저장될 수도 있다.
- [444] 프로그램 모듈들은 일 실시예에 따른 기능 또는 동작을 수행하거나, 일 실시예에 따른 추상 데이터 유형을 구현하는 루틴(routine), 서브루틴(subroutine), 프로그램, 오브젝트(object), 컴퍼넌트(component) 및 데이터 구조(data structure) 등을 포괄할 수 있지만, 이에 제한되지는 않는다.
- [445] 프로그램 모듈들은 복호화 장치(2000)의 적어도 하나의 프로세서(processor)에 의해 수행되는 명령어(instruction) 또는 코드(code)로 구성될 수 있다.
- [446] 처리부(2010)는 전술된 복호화기에 대응할 수 있다. 말하자면, 도 18을 참조하여 전술된 복호화기 및 도 9를 참조하여 전술된 자동 부호기의 복호화에 대한 동작은 처리부(2010)에 의해 수행될 수 있다.
- [447] 저장부는 메모리(2030) 및/또는 저장소(2040)를 나타낼 수 있다. 메모리(2030) 및 저장소(2040)는 다양한 형태의 휘발성 또는 비휘발성 저장 매체일 수 있다. 예를 들면, 메모리(2030)는 롬(ROM)(2031) 및 램(RAM)(2032) 중 적어도 하나를 포함할 수 있다.
- [448] 저장부는 복호화 장치(2000)의 동작을 위해 사용되는 데이터 또는 정보를 저장할 수 있다. 실시예에서, 복호화 장치(2000)가 갖는 데이터 또는 정보는 저장부 내에 저장될 수 있다.
- [449] 복호화 장치(2000)는 컴퓨터에 의해 독출(read)될 수 있는 기록 매체를 포함하는 컴퓨터 시스템에서 구현될 수 있다.
- [450] 기록 매체는 복호화 장치(2000)가 동작하기 위해 요구되는 적어도 하나의 모듈을 저장할 수 있다. 메모리(2030)는 적어도 하나의 모듈을 저장할 수 있고, 적어도 하나의 모듈이 처리부(2010)에 의하여 실행되도록 구성될 수 있다.
- [451] 복호화 장치(2000)의 데이터 또는 정보의 통신과 관련된 기능은 통신부(2020)를 통해 수행될 수 있다.
- [452] 네트워크(2099)는 부호화 장치(1900) 및 복호화 장치(2000) 간의 통신을 제공할 수 있다.
- [453] 도 21는 일 실시예에 따른 부호화 방법의 흐름도이다.
- [454] 단계(2110)에서, 부호화 장치(1900)의 처리부(1910)는 비트스트림을 생성할 수 있다.
- [455] 처리부(1910)는 입력 이미지에 대해 엔트로피 모델을 사용하는 엔트로피 부호화를 수행하여 비트스트림을 생성할 수 있다.

- [456] 처리부(1910)는 도 17을 참조하여 전송된 부호화기 및 도 9를 참조하여 전송된 자동 부호기의 부호화에 대한 동작을 수행할 수 있다. 처리부(1910)는 부호화에 있어서 이미지 압축 네트워크 및 품질 향상 네트워크를 사용할 수 있다.
- [457] 단계(2120)에서, 부호화 장치(1900)의 통신부(1920)는 비트스트림을 전송할 수 있다. 통신부(1920)는 비트스트림을 복호화 장치(2000)로 전송할 수 있다. 또는, 비트스트림은 부호화 장치(1900)의 저장부에 저장될 수 있다.
- [458] 전송된 실시예에서 설명된 이미지의 엔트로피 부호화 및 엔트로피 엔진에 관련된 내용은 본 실시예에도 적용될 수 있다. 중복되는 설명은 생략된다.
- [459] 도 22는 일 실시예에 따른 복호화 방법의 흐름도이다.
- [460] 단계(2210)에서, 복호화 장치(2000)의 통신부(2020) 또는 저장부는 비트스트림을 획득할 수 있다.
- [461] 단계(2220)에서, 복호화 장치(2000)의 처리부(2010)는 비트스트림을 사용하여 재구성된 이미지를 생성할 수 있다.
- [462] 복호화 장치(2000)의 처리부(2010)는 비트스트림에 대해 엔트로피 모델을 사용하는 복호화를 수행하여 재구성된 이미지를 생성할 수 있다.
- [463] 처리부(2010)는 도 18을 참조하여 전송된 복호화기 및 도 9를 참조하여 전송된 자동 부호기의 복호화에 대한 동작을 수행할 수 있다.
- [464] 처리부(2010)는 복호화에 있어서 이미지 압축 네트워크 및 품질 향상 네트워크를 사용할 수 있다.
- [465] 전송된 실시예에서 설명된 이미지의 엔트로피 복호화 및 엔트로피 엔진에 관련된 내용은 본 실시예에도 적용될 수 있다. 중복되는 설명은 생략된다.
- [466] 이미지에 대한 패딩
- [467] 도 23은 일 예에 따른 입력 이미지로의 패딩을 나타낸다.
- [468] 도 23에서는, 입력 이미지의 중심부로의 패딩을 통해, 입력 이미지의 크기가  $w \times y$ 로부터  $w + pw \times h + ph$ 로 변하는 것이 도시되었다.
- [469] 고수준의 MS-SSIM을 획득하기 위해 패딩 방법이 사용될 수 있다.
- [470] 실시예의 이미지 압축 방법에서는,  $y$  생성 및  $z$  생성의 단계에서 1/2의 다운-스케일링이 수행될 수 있다. 따라서, 입력 이미지의 크기가  $2^n$ 의 배수일 경우에, 최대의 압축 성능이 도출될 수 있다. 여기에서,  $n$ 은 입력 이미지에 대한 다운-스케일링의 개수일 수 있다.
- [471] 예를 들면, 도 9를 참조하여 전송된 실시예에서는  $x$ 로부터  $y$ 로의 1/2 다운-스케일링이 4회 수행될 수 있고,  $y$ 로부터  $z$ 로의 1/2 다운-스케일링이 2회 수행될 수 있다. 따라서, 입력 이미지의 크기는  $2^6 (= 64)$ 의 배수가 되는 것이 바람직할 수 있다.
- [472] 또한, 패딩의 위치와 관련하여, MS-SSIM과 같은 특정된 방식이 사용되는 경우,

입력 이미지의 경계에 대한 패딩보다, 입력 이미지의 중심부에 패딩이 이루어지는 것이 더 바람직하다.

[473] 도 24는 일 실시예에 따른 부호화에서의 패딩을 위한 코드를 나타낸다.

[474] 도 25는 일 실시예에 따른 부호화에서의 패딩 방법의 흐름도이다.

[475] 도 21을 참조하여 전송된 단계(2110)는 단계들(2510, 2520, 2530 및 2540)을 포함할 수 있다.

[476] 이하에서, 기준 값  $k$ 는  $2^n$ 일 수 있다.  $n$ 은 이미지 압축 네트워크에서의 입력 이미지에 대한 다운-스케일링들의 개수일 수 있다.

[477] 단계(2510)에서, 처리부(1910)는 입력 이미지에 수평 방향의 패딩을 적용할지 여부를 판단할 수 있다.

[478] 수평 방향의 패딩은 입력 이미지의 수직 축 상의 중심에 하나 이상의 행들을 삽입하는 것일 수 있다.

[479] 예를 들면, 처리부(1910)는 입력 이미지의 높이  $h$  및 기준 값  $k$ 에 기반하여 입력 이미지에 수평 방향의 패딩을 적용할지 여부를 판단할 수 있다. 처리부(1910)는 입력 이미지의 높이  $h$ 가 기준 값  $k$ 의 배수가 아니면 입력 이미지에 수평 방향의 패딩을 적용할 수 있다. 처리부(1910)는 입력 이미지의 높이  $h$ 가 기준 값  $k$ 의 배수이면 입력 이미지에 수평 방향의 패딩을 적용하지 않을 수 있다.

[480] 입력 이미지에 수평 방향의 패딩을 적용하는 경우 단계(2520)가 수행될 수 있다.

[481] 입력 이미지에 수평 방향의 패딩을 적용하지 않는 경우 단계(2530)가 수행될 수 있다.

[482] 단계(2520)에서, 처리부(1910)는 입력 이미지에 수평 방향의 패딩을 적용할 수 있다. 처리부(1910)는 입력 이미지의 상측 영역 및 입력 이미지의 하측 영역 사이에 패딩 영역을 추가할 수 있다.

[483] 처리부(1910)는 입력 이미지에 수평 방향의 패딩을 적용함으로써 입력 이미지의 높이를 기준 값  $k$ 의 배수가 되도록 조정할 수 있다.

[484] 예를 들면, 처리부(1910)는 입력 이미지를 수직 방향으로 분리함으로써 상단 이미지 및 하단 이미지를 생성할 수 있다. 처리부(1910)는 상단 이미지 및 하단 이미지의 사이에 패딩을 적용할 수 있다. 처리부(1910)는 패딩 영역을 생성할 수 있다. 처리부(1910)는 상단 이미지, 패딩 영역 및 하단 이미지를 결합함으로써 높이가 조절된 입력 이미지를 생성할 수 있다.

[485] 여기에서, 패딩은 모서리(edge) 패딩일 수 있다.

[486] 단계(2530)에서, 처리부(1910)는 입력 이미지에 수직 방향의 패딩을 적용할지 여부를 판단할 수 있다.

[487] 수직 방향의 패딩은 입력 이미지의 수평 축 상의 중심에 하나 이상의 열들을 삽입하는 것일 수 있다.

[488] 예를 들면, 처리부(1910)는 입력 이미지의 넓이  $w$  및 기준 값  $k$ 에 기반하여 입력 이미지에 수직 방향의 패딩을 적용할지 여부를 판단할 수 있다. 처리부(1910)는

- 입력 이미지의 넓이  $w$ 가 기준 값  $k$ 의 배수가 아니면 입력 이미지에 수직 방향의 패딩을 적용할 수 있다. 처리부(1910)는 입력 이미지의 넓이  $w$ 가 기준 값  $k$ 의 배수이면 입력 이미지에 수직 방향의 패딩을 적용하지 않을 수 있다.
- [489] 입력 이미지에 수직 방향의 패딩을 적용하는 경우 단계(2540)가 수행될 수 있다.
- [490] 입력 이미지에 수직 방향의 패딩을 적용하지 않는 경우 절차가 종료할 수 있다.
- [491] 단계(2540)에서, 처리부(1910)는 입력 이미지에 수직 방향의 패딩을 적용할 수 있다. 처리부(1910)는 입력 이미지의 좌측 영역 및 입력 이미지의 우측 영역 사이에 패딩 영역을 추가할 수 있다.
- [492] 처리부(1910)는 입력 이미지에 수직 방향의 패딩을 적용함으로써 입력 이미지의 폭을 기준 값  $k$ 의 배수가 되도록 조정할 수 있다.
- [493] 예를 들면, 처리부(1910)는 입력 이미지를 수직 방향으로 분리함으로써 좌측 이미지 및 우측 이미지를 생성할 수 있다. 처리부(1910)는 좌측 이미지 및 우측 이미지의 사이에 패딩을 적용할 수 있다. 처리부(1910)는 패딩 영역을 생성할 수 있다. 처리부(1910)는 좌측 이미지, 패딩 영역 및 우측 이미지를 결합함으로써 폭이 조절된 입력 이미지를 생성할 수 있다.
- [494] 여기에서, 패딩은 모서리(edge) 패딩일 수 있다.
- [495] 전술된 단계들(2510, 2520, 2530 및 2540)의 패딩을 통해 패딩된 이미지가 생성될 수 있다. 패딩된 이미지의 폭 및 높이는 각각 기준 값  $k$ 의 배수일 수 있다.
- [496] 패딩된 이미지는 입력 이미지를 대체하여 사용될 수 있다.
- [497] 도 26은 일 실시예에 따른 부호화에서의 패딩 영역의 제거를 위한 코드를 나타낸다.
- [498] 도 27은 일 실시예에 따른 부호화에서의 패딩의 제거 방법의 흐름도이다.
- [499] 도 22를 참조하여 전술된 단계(2220)는 단계들(2710, 2720, 2730 및 2740)을 포함할 수 있다.
- [500] 이하에서, 대상 이미지는 도 25를 참조하여 전술된 실시예의 패딩이 적용된 이미지에 대해 재구축된 이미지일 수 있다. 말하자면, 대상 이미지는 입력 이미지에 대한, 패딩, 부호화 및 복호화를 통해 생성된 이미지일 수 있다. 이하에서, 원 이미지의 높이  $h$ 는 수평 방향의 패딩이 적용되기 전의 입력 이미지의 높이를 의미할 수 있다. 원 이미지의 폭  $w$ 는 수직 방향의 패딩이 적용되기 전의 입력 이미지의 폭을 의미할 수 있다.
- [501] 이하에서, 기준 값  $k$ 는  $2^n$ 일 수 있다.  $n$ 은 이미지 압축 네트워크에서의 입력 이미지에 대한 다운-스케일링들의 개수일 수 있다.
- [502] 단계(2710)에서, 처리부(2010)는 대상 이미지로부터 수평 방향의 패딩 영역을 제거할지 여부를 판단할 수 있다.
- [503] 수평 방향의 패딩 영역의 제거는 대상 이미지의 수직 축 상의 중심에서 하나 이상의 행들을 제거하는 것일 수 있다.
- [504] 예를 들면, 처리부(2010)는 원 이미지의 높이  $h$  및 기준 값  $k$ 에 기반하여 대상

이미지로부터 수평 방향의 패딩 영역을 제거할지 여부를 판단할 수 있다. 처리부(2010)는 원 이미지의 높이  $h$ 가 기준 값  $k$ 의 배수가 아니면 대상 이미지로부터 수평 방향의 패딩 영역을 제거할 수 있다. 처리부(2010)는 원 이미지의 높이  $h$ 가 기준 값  $k$ 의 배수이면 대상 이미지로부터 수평 방향의 패딩 영역을 제거하지 않을 수 있다.

- [505] 예를 들면, 처리부(2010)는 원 이미지의 높이  $h$  및 대상 이미지의 높이에 기반하여 이미지로부터 대상 이미지로부터 수평 방향의 패딩 영역을 제거할지 여부를 판단할 수 있다. 처리부(2010)는 원 이미지의 높이  $h$  및 대상 이미지의 높이가 동일하지 않으면 대상 이미지로부터 수평 방향의 패딩 영역을 제거할 수 있다. 처리부(2010)는 원 이미지의 높이  $h$  및 대상 이미지의 높이가 동일하면 대상 이미지로부터 수평 방향의 패딩 영역을 제거하지 않을 수 있다.
- [506] 대상 이미지로부터 수평 방향의 패딩 영역을 제거하는 경우 단계(2720)가 수행될 수 있다.
- [507] 대상 이미지로부터 수평 방향의 패딩 영역을 제거하지 않는 경우 단계(2730)가 수행될 수 있다.
- [508] 단계(2720)에서, 처리부(2010)는 대상 이미지로부터 수평 방향의 패딩 영역을 제거할 수 있다. 처리부(2010)는 대상 이미지의 상측 영역 및 입력 이미지의 하측 영역 사이의 패딩 영역을 제거할 수 있다.
- [509] 예를 들면, 처리부(2010)는 대상 이미지로부터 수평 방향의 패딩 영역을 제거함으로써 상단 이미지 및 하단 이미지를 생성할 수 있다. 처리부(2010)는 상단 이미지 및 하단 이미지를 결합함으로써 대상 이미지의 높이를 조절할 수 있다.
- [510] 패딩 영역의 제거를 통해 대상 이미지의 높이가 원 이미지의 높이  $h$ 와 동일하게 될 수 있다.
- [511] 여기에서, 패딩 영역은 모서리(edge) 패딩에 의해 생성된 영역일 수 있다.
- [512] 단계(2730)에서, 처리부(2010)는 대상 이미지로부터 수직 방향의 패딩 영역을 제거할지 여부를 판단할 수 있다.
- [513] 수직 방향의 패딩 영역의 제거는 대상 이미지의 수평 축 상의 중심에서 하나 이상의 열들을 제거하는 것일 수 있다.
- [514] 예를 들면, 처리부(2010)는 원 이미지의 넓이  $w$  및 기준 값  $k$ 에 기반하여 대상 이미지로부터 수직 방향의 패딩 영역을 제거할지 여부를 판단할 수 있다. 처리부(2010)는 원 이미지의 넓이  $w$ 가 기준 값  $k$ 의 배수가 아니면 대상 이미지로부터 수직 방향의 패딩 영역을 제거할 수 있다. 처리부(2010)는 원 이미지의 넓이  $w$ 가 기준 값  $k$ 의 배수이면 대상 이미지로부터 수직 방향의 패딩 영역을 제거하지 않을 수 있다.
- [515] 예를 들면, 처리부(2010)는 원 이미지의 넓이  $w$  및 대상 이미지의 넓이에 기반하여 이미지로부터 대상 이미지로부터 수직 방향의 패딩 영역을 제거할지 여부를 판단할 수 있다. 처리부(2010)는 원 이미지의 넓이  $w$  및 대상 이미지의

넓이가 동일하지 않으면 대상 이미지로부터 수직 방향의 패딩 영역을 제거할 수 있다. 처리부(2010)는 원 이미지의 넓이  $w$  및 대상 이미지의 넓이가 동일하면 대상 이미지로부터 수직 방향의 패딩 영역을 제거하지 않을 수 있다.

- [516] 대상 이미지로부터 수직 방향의 패딩 영역을 제거하는 경우 단계(2740)가 수행될 수 있다.
- [517] 대상 이미지로부터 수직 방향의 패딩 영역을 제거하지 않는 경우 절차가 종료할 수 있다.
- [518] 단계(2740)에서, 처리부(2010)는 대상 이미지로부터 수직 방향의 패딩 영역을 제거할 수 있다. 처리부(2010)는 대상 이미지의 좌측 영역 및 입력 이미지의 우측 영역 사이의 패딩 영역을 제거할 수 있다.
- [519] 예를 들면, 처리부(2010)는 대상 이미지로부터 수직 방향의 패딩 영역을 제거함으로써 좌측 이미지 및 우측 이미지를 생성할 수 있다. 처리부(2010)는 좌측 이미지 및 우측 이미지를 결합함으로써 대상 이미지의 폭을 조절할 수 있다.
- [520] 여기에서, 패딩 영역은 모서리(edge) 패딩에 의해 생성된 영역일 수 있다.
- [521] 전술된 단계들(2710, 2720, 2730 및 2740)에 의해 대상 이미지로부터 패딩이 제거될 수 있다.
- [522] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [523] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는

데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

- [524] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다.
- [525] 컴퓨터 판독 가능한 기록 매체는 본 발명에 따른 실시예들에서 사용되는 정보를 포함할 수 있다. 예를 들면, 컴퓨터 판독 가능한 기록 매체는 비트스트림을 포함할 수 있고, 비트스트림은 본 발명에 따른 실시예들에서 설명된 정보를 포함할 수 있다.
- [526] 컴퓨터 판독 가능한 기록 매체는 비-일시적 컴퓨터 판독 가능한 매체(non-transitory computer-readable medium)를 포함할 수 있다.
- [527] 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [528] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [529] 실시예에서 설명된 장치는 하나 이상의 프로세서들을 포함할 수 있고, 메모리를 포함할 수 있다. 메모리는 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장할 수 있다. 하나 이상의 프로그램들은 실시예에서 설명된 장치의 동작을 수행할 수 있다. 예를 들면, 장치의 하나

이상의 프로그램들은 전술된 단계들 중 장치와 관련된 단계에서 설명된 동작을 수행할 수 있다. 말하자면, 실시예에서 설명된 장치의 동작은 하나 이상의 프로그램들에 의해 실행될 수 있다. 하나 이상의 프로그램들은 실시예에서 전술된 장치의 프로그램, 어플리케이션 및 앱 등을 포함할 수 있다. 예를 들면, 하나 이상의 프로그램들 중 하나의 프로그램은 실시예에서 전술된 장치의 프로그램, 어플리케이션 및 앱에 대응할 수 있다.

## 청구범위

- [청구항 1] 입력 이미지에 대해 엔트로피 모델을 사용하는 엔트로피 부호화를 수행하여 비트스트림을 생성하는 단계; 및 상기 비트스트림을 전송 또는 저장하는 단계를 포함하는 부호화 방법.
- [청구항 2] 제1항에 있어서, 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델이고, 상기 문맥-적응형 엔트로피 모델은 문맥들의 서로 상이한 3 개의 타입들을 활용하는 부호화 방법.
- [청구항 3] 제2항에 있어서, 상기 문맥들은 가우시안 혼합 모델의 파라미터를 추정하기 위해 사용되는 부호화 방법.
- [청구항 4] 제3항에 있어서, 상기 파라미터는 가중치 파라미터, 평균 파라미터 및 표준 편차 파라미터를 포함하는 부호화 방법.
- [청구항 5] 제1항에 있어서, 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델이고, 상기 문맥-적응형 엔트로피 모델은 전역 문맥을 사용하는 부호화 방법.
- [청구항 6] 제1항에 있어서, 상기 엔트로피 부호화는 이미지 압축 네트워크 및 품질 향상 네트워크의 결합에 의해 수행되는 부호화 방법.
- [청구항 7] 제6항에 있어서, 상기 품질 향상 네트워크는 매우 깊은 슈퍼 레졸루션(Very Deep Super Resolution; VDSR), 잔차 밀도 네트워크(Residual Dense Network; RDN) 또는 그룹된 잔차 밀도 네트워크(Grouped Residual Dense Network; GRDN)인 부호화 방법.
- [청구항 8] 제1항에 있어서, 상기 입력 이미지에 수평 방향의 패딩 또는 수직 방향의 패딩이 적용되고, 상기 수평 방향의 패딩은 상기 입력 이미지의 수직 축 상의 중심에 하나 이상의 행들을 삽입하는 것이고, 상기 수직 방향의 패딩은 상기 입력 이미지의 수평 축 상의 중심에 하나 이상의 열들을 삽입하는 것인 부호화 방법.
- [청구항 9] 제8항에 있어서, 상기 수평 방향의 패딩은 상기 입력 이미지의 높이가  $k$ 의 배수가 아닐 경우에 수행되고, 상기 수직 방향의 패딩은 상기 입력 이미지의 폭이  $k$ 의 배수가 아닐 경우에 수행되고,

- 상기  $k$ 는  $2^n$ 이고,  
 상기  $n$ 은 상기 입력 이미지에 대한 다운-스케일링들의 개수인 부호화 방법.
- [청구항 10] 제1항에 기재된 부호화 방법에 의하여 생성된 상기 비트스트림을 기록하는 기록 매체.
- [청구항 11] 비트스트림을 획득하는 통신부; 및  
 상기 비트스트림에 대해 엔트로피 모델을 사용하는 복호화를 수행하여 재구성된 이미지를 생성하는 처리부를 포함하는 복호화 장치.
- [청구항 12] 비트스트림을 획득하는 단계; 및  
 상기 비트스트림에 대해 엔트로피 모델을 사용하는 복호화를 수행하여 재구성된 이미지를 생성하는 단계를 포함하는 복호화 방법.
- [청구항 13] 제12항에 있어서,  
 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델이고,  
 상기 문맥-적응형 엔트로피 모델은 문맥들의 서로 상이한 3 개의 타입들을 활용하는 복호화 방법.
- [청구항 14] 제13항에 있어서,  
 상기 문맥들은 가우시안 혼합 모델의 파라미터를 추정하기 위해 사용되는 복호화 방법.
- [청구항 15] 제14항에 있어서,  
 상기 파라미터는 가중치 파라미터, 평균 파라미터 및 표준 편차 파라미터를 포함하는 복호화 방법.
- [청구항 16] 제12항에 있어서,  
 상기 엔트로피 모델은 문맥-적응형 엔트로피 모델이고,  
 상기 문맥-적응형 엔트로피 모델은 전역 문맥을 사용하는 복호화 방법.
- [청구항 17] 제12항에 있어서,  
 상기 엔트로피 부호화는 이미지 압축 네트워크 및 품질 향상 네트워크의 결합에 의해 수행되는 복호화 방법.
- [청구항 18] 제12항에 있어서,  
 상기 품질 향상 네트워크는 매우 깊은 슈퍼 레졸루션(Very Deep Super Resolution; VDSR), 잔차 밀도 네트워크(Residual Dense Network; RDN) 또는 그룹된 잔차 밀도 네트워크(Grouped Residual Dense Network; GRDN)인 복호화 방법.
- [청구항 19] 제12항에 있어서,  
 상기 재구성된 이미지로부터 수평 방향의 패딩 영역 또는 수직 방향의 패딩 영역이 제거되고,  
 상기 수평 방향의 패딩 영역의 제거는 상기 재구성된 이미지의 수직 축

상의 중심에서 하나 이상의 행들을 제거하는 것이고,  
 상기 수직 방향의 패딩 영역의 제거는 상기 재구축된 이미지의 수평 축  
 상의 중심에서 하나 이상의 열들을 제거하는 것인 복호화 방법.

[청구항 20]

제19항에 있어서,

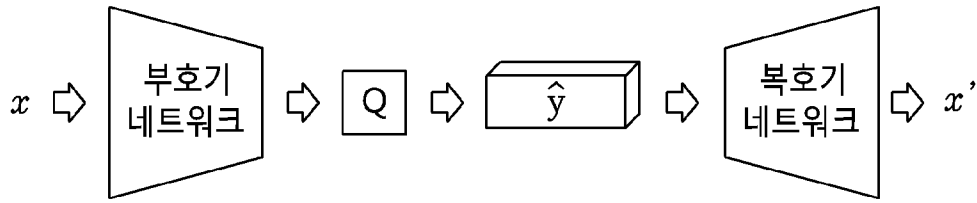
상기 수평 방향의 패딩 영역의 제거는 원 이미지의 높이가  $k$ 의 배수가  
 아닐 경우에 수행되고,

상기 수직 방향의 패딩 영역의 제거는 상기 원 이미지의 폭이  $k$ 의 배수가  
 아닐 경우에 수행되고,

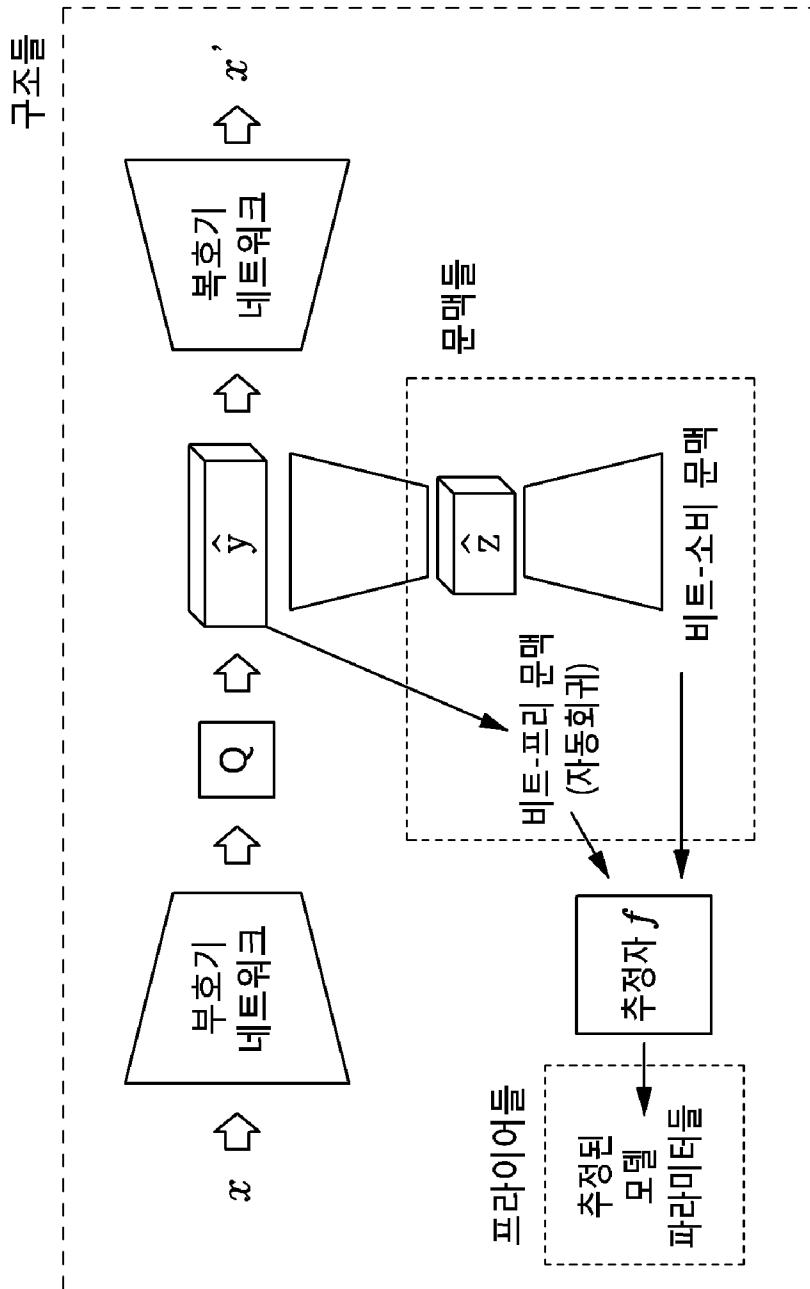
상기  $k$ 는  $2^n$ 이고,

상기  $n$ 은 상기 원 이미지에 대한 다운-스케일링들의 개수인 복호화 방법.

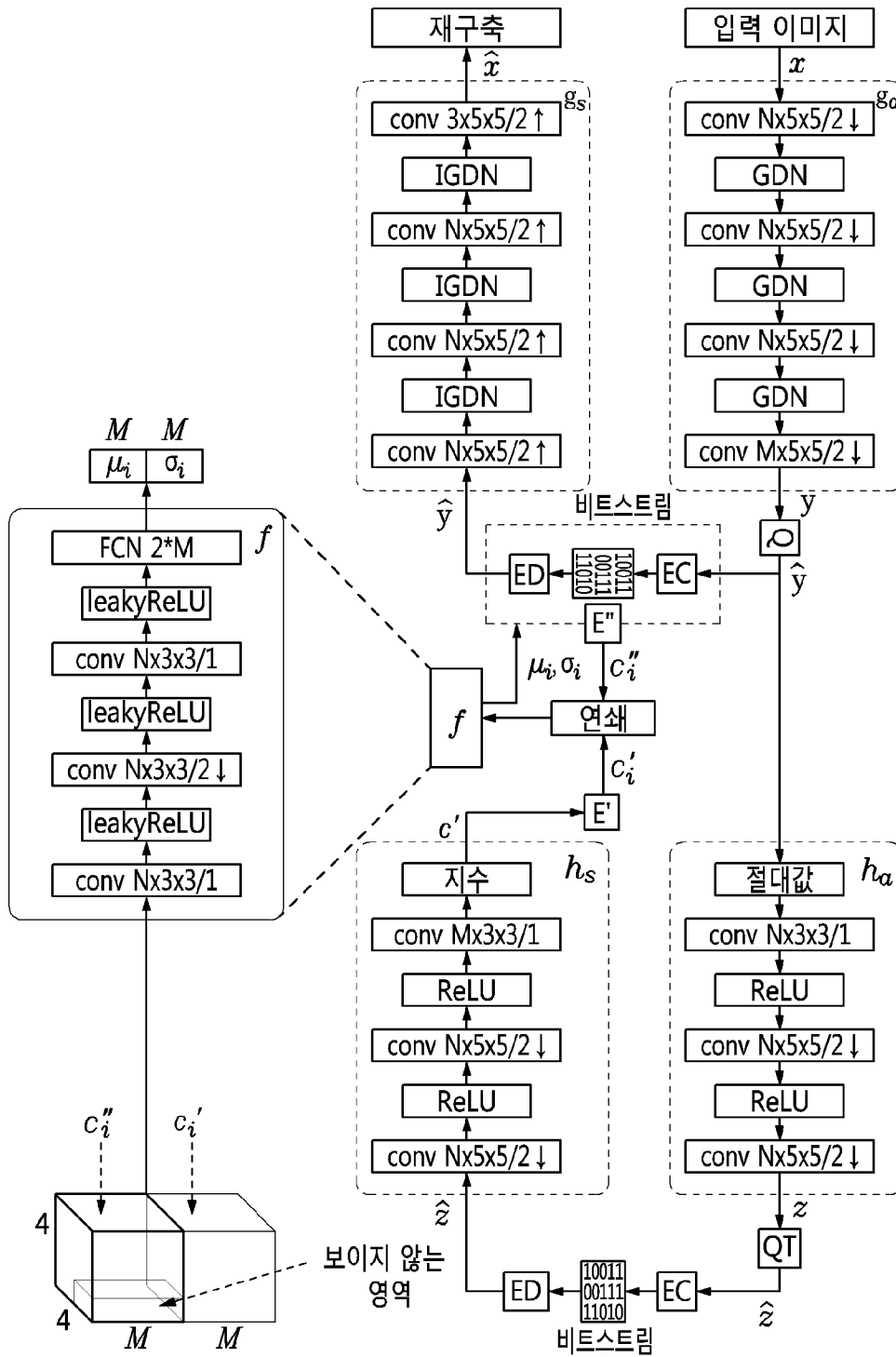
[도1]



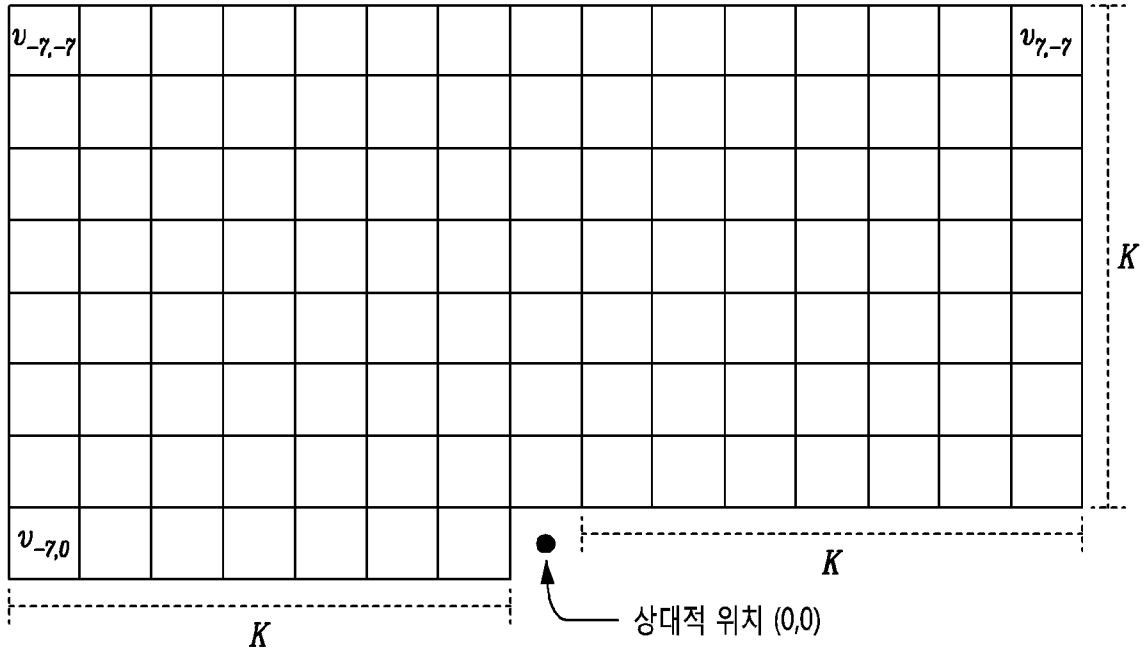
[도2]



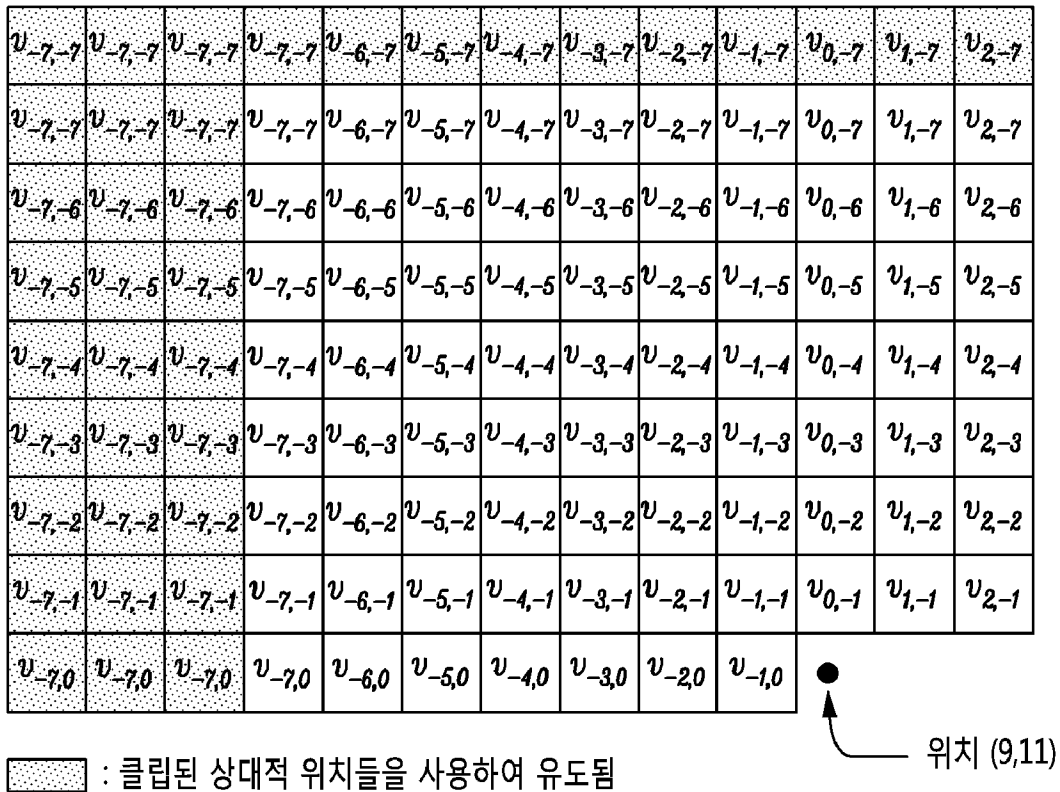
[도3]



[도4]



[도5]



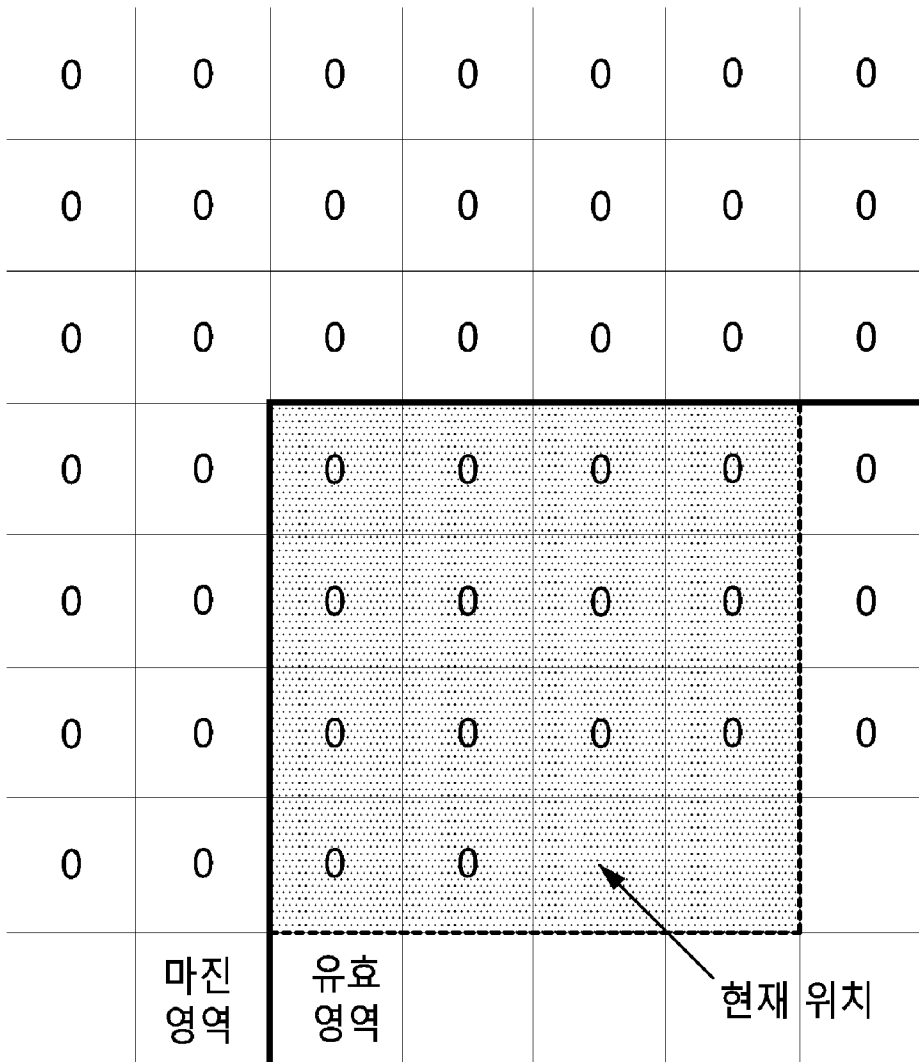
[도6]

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0					
0	0					
0	0					
	마진 영역	유효 영역				

현재 위치

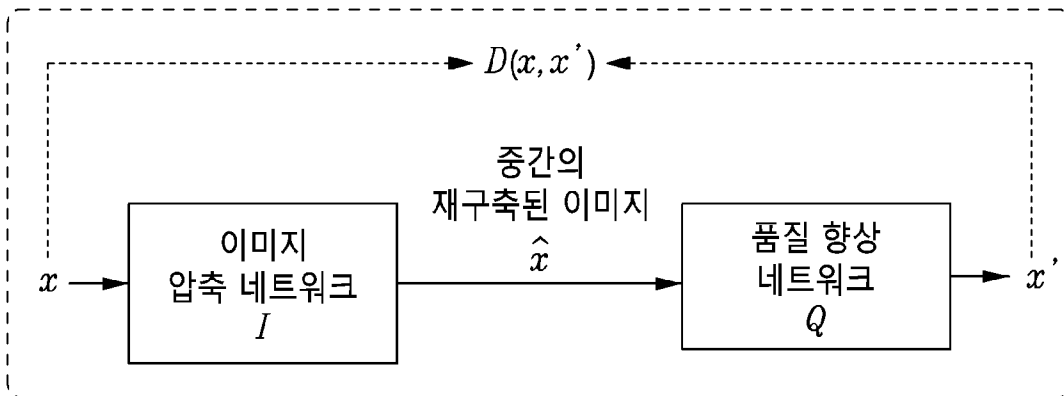
The diagram shows a grid of 7 rows and 7 columns. The top 4 rows and the first 4 columns are shaded with a dotted pattern. A thick black L-shaped line is drawn at the intersection of the 4th row and 4th column. An arrow points from the text '현재 위치' (Current Position) to the cell at the intersection of the 4th row and 4th column. The bottom row contains the labels '마진 영역' (Margin Area) under the first two columns and '유효 영역' (Valid Area) under the third column. The cells containing '0' are located in the top 4 rows and the first two columns of the bottom 3 rows.

[도7]

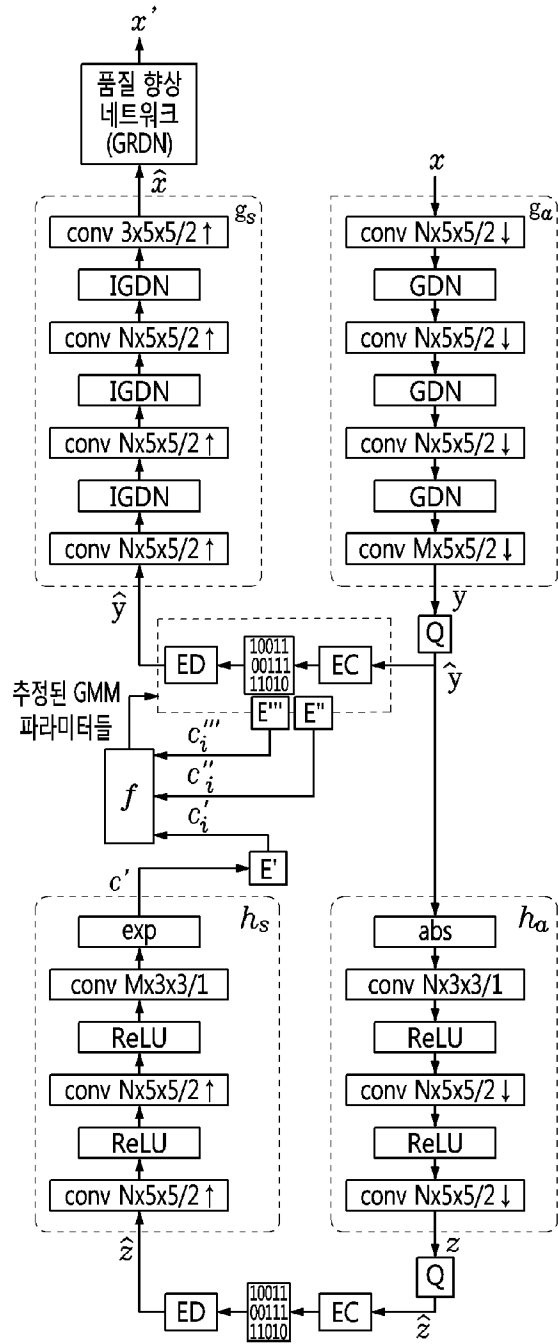


[도8]

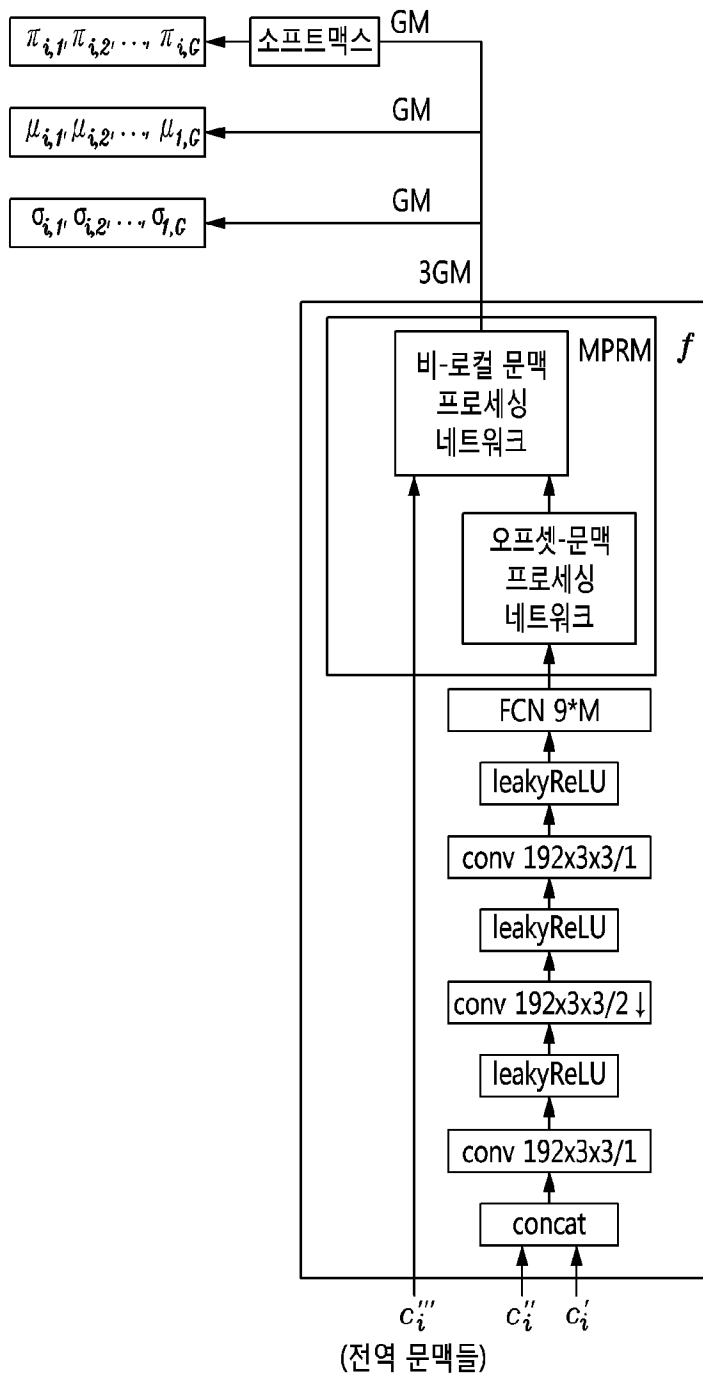
학습된 이미지 압축 프레임워크



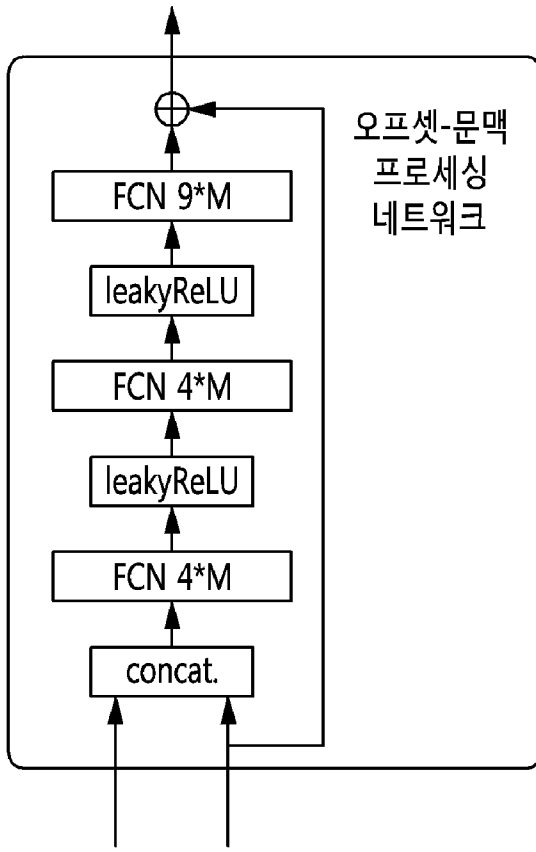
[도9]



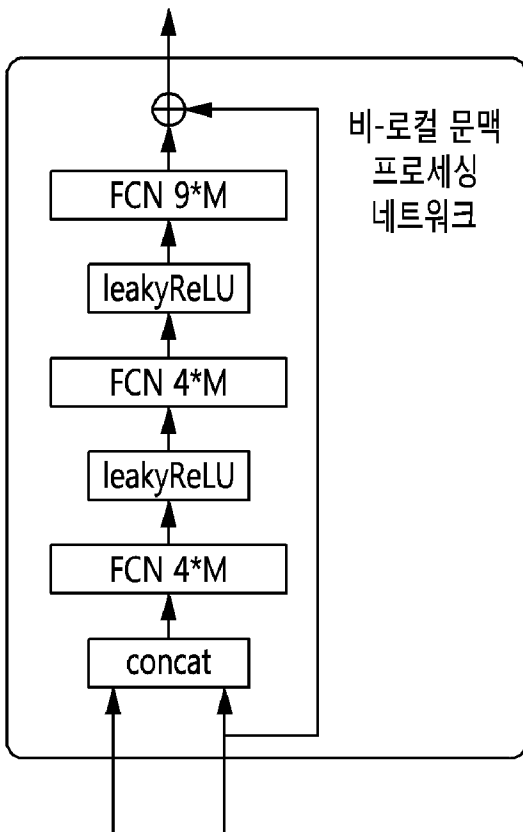
[도 10]



[도11]

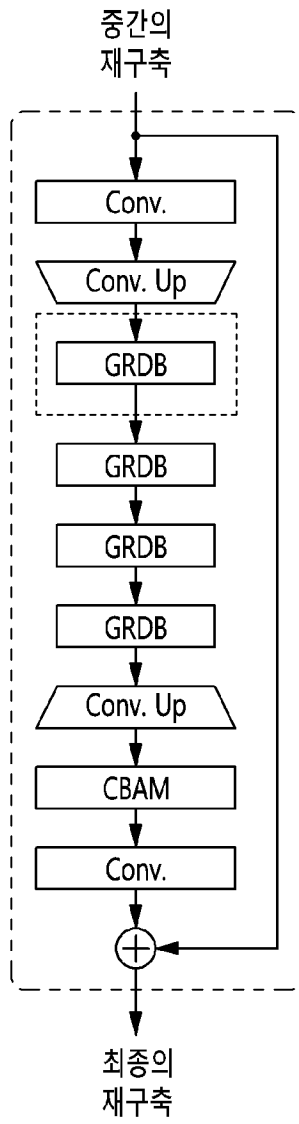


[도12]

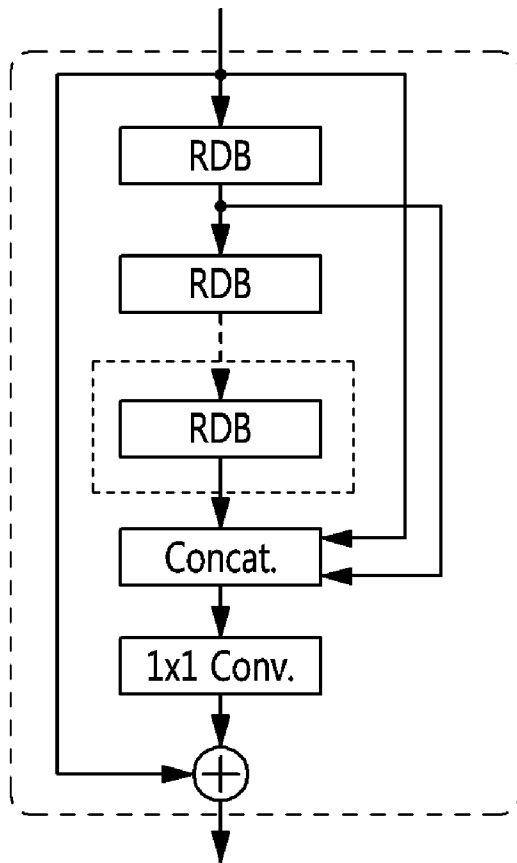




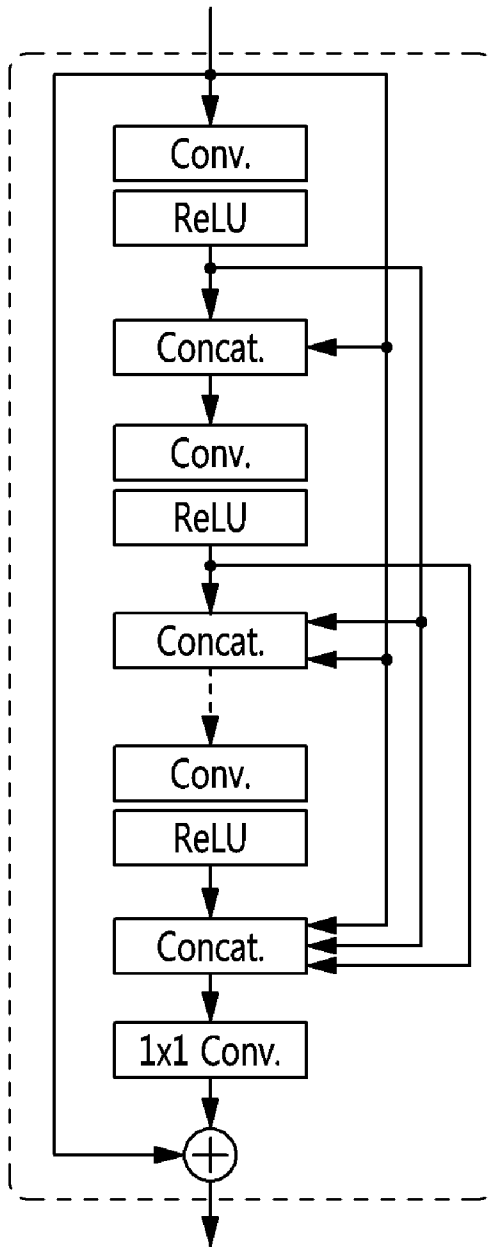
[도 14]



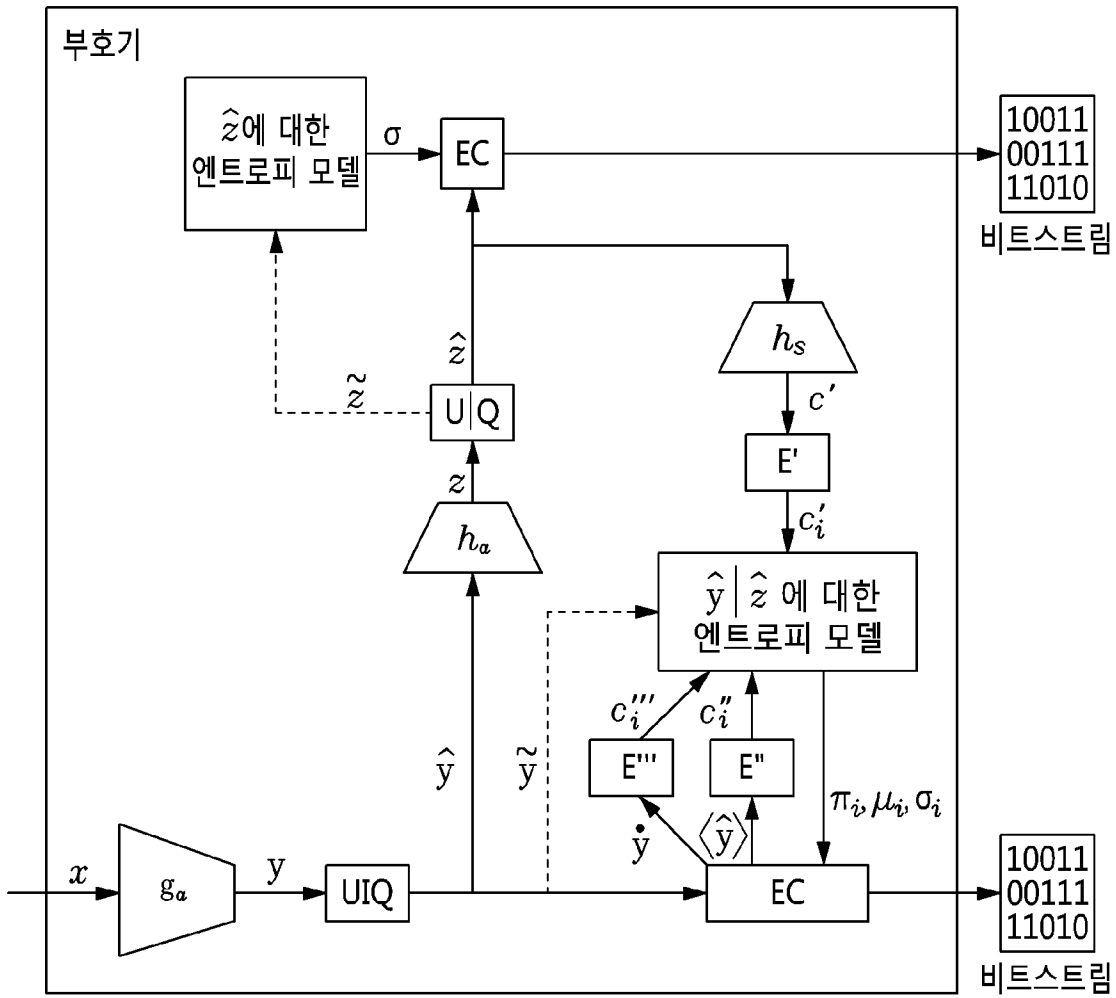
[도 15]



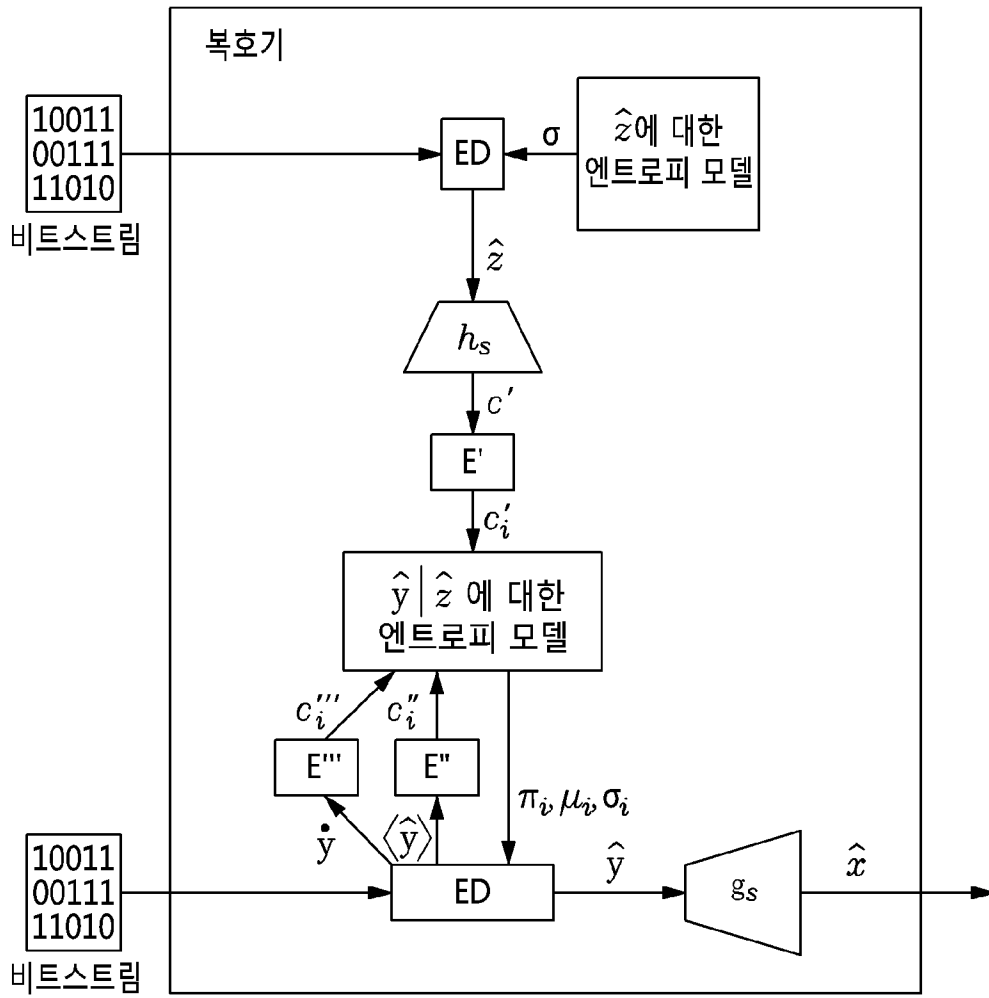
[도 16]



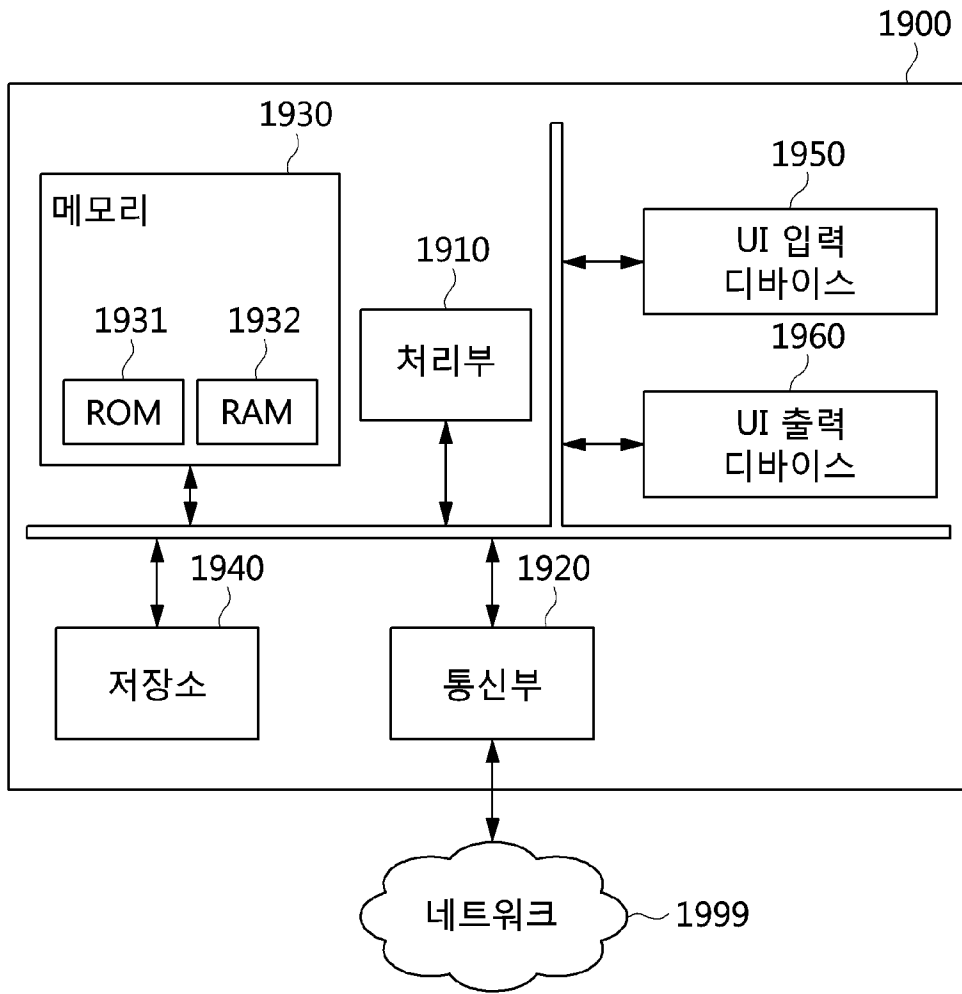
[도17]



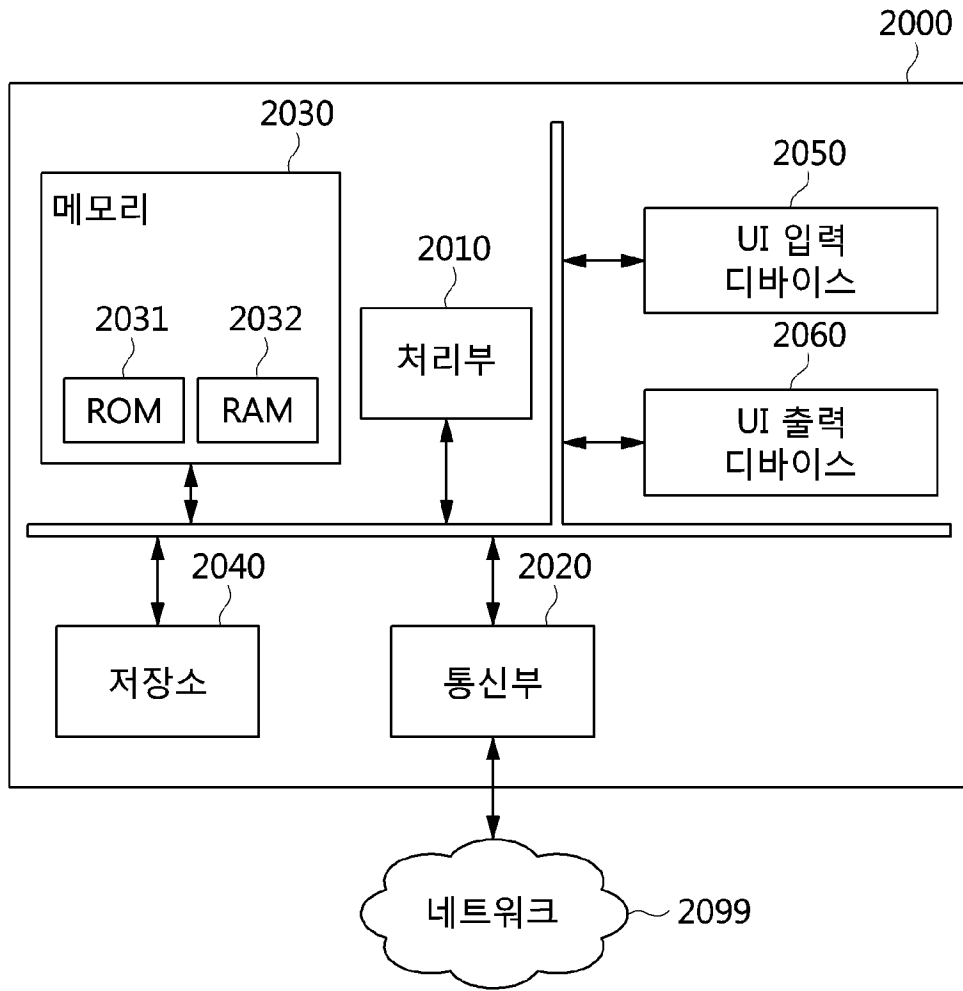
[도 18]



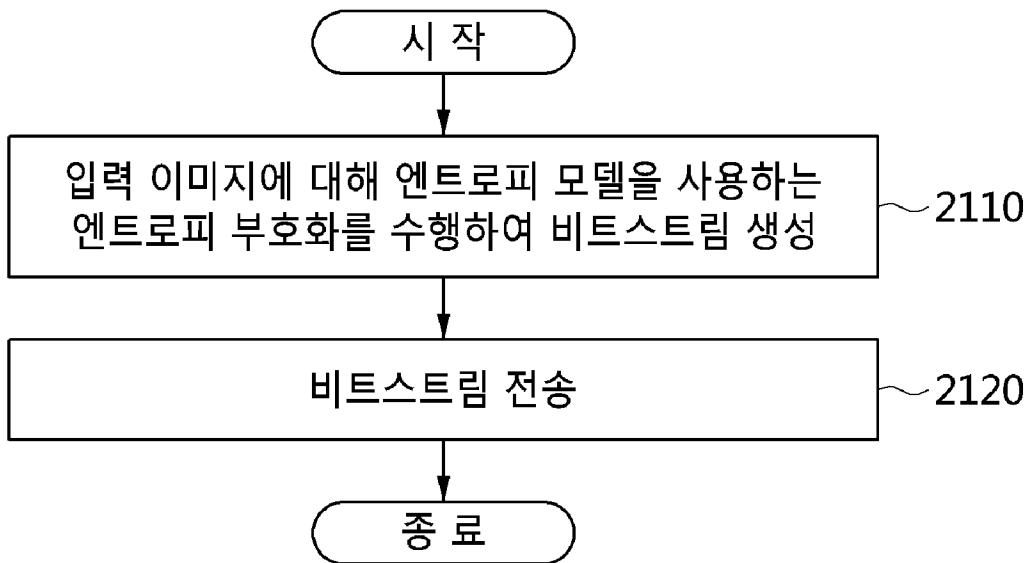
[도 19]



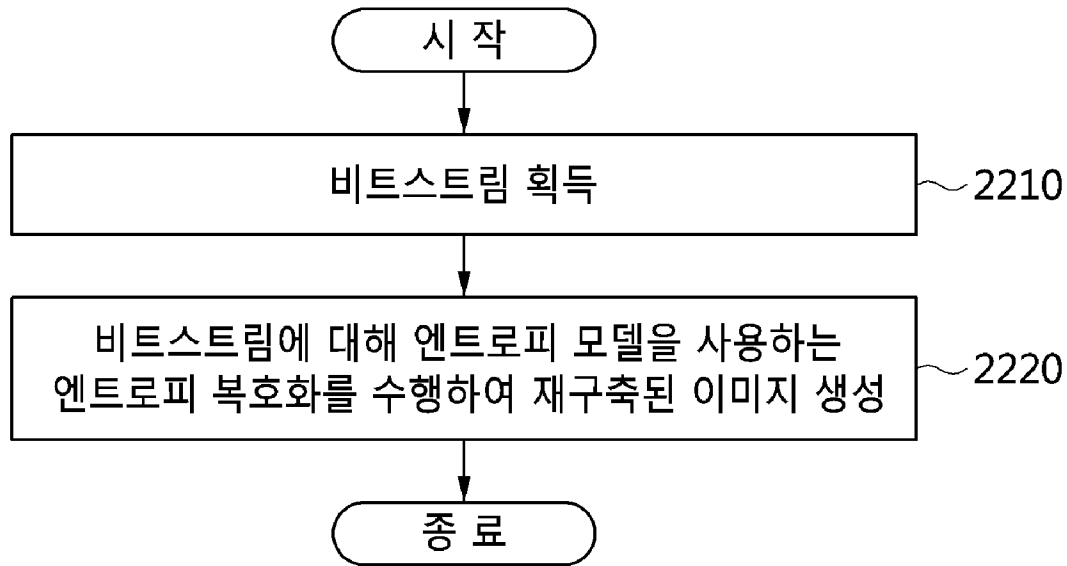
[도20]



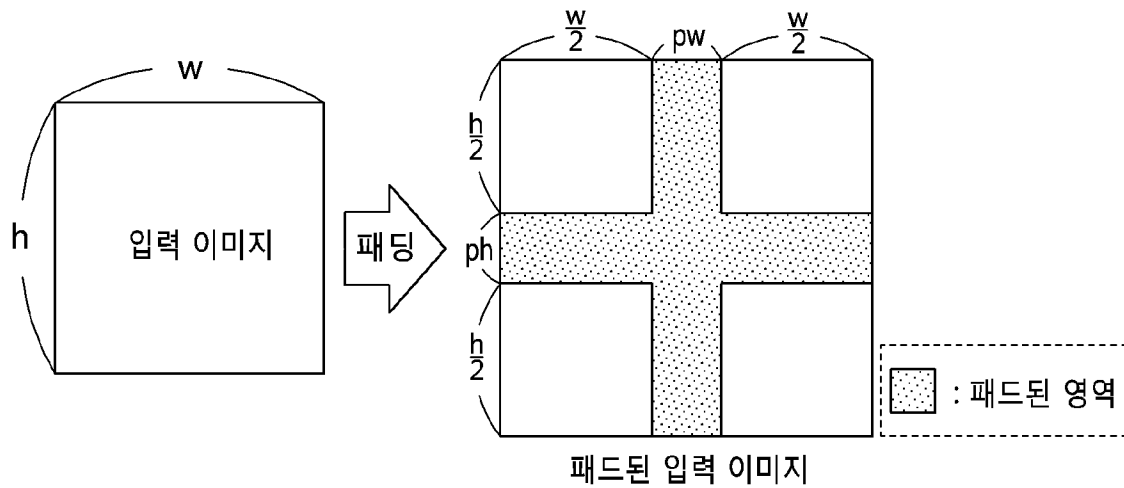
[도21]



[도22]



[도23]



[도24]

```
img = Image.open(input_path)
w, h = img.size

new_w = int(math.ceil(w / K) * K)
new_h = int(math.ceil(h / K) * K)

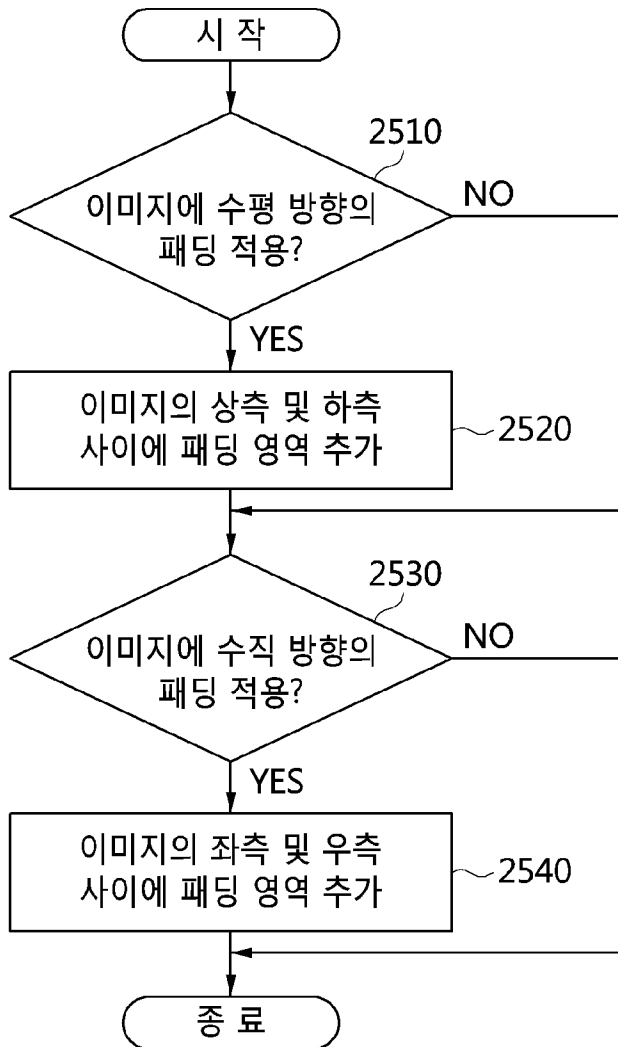
h_split_offset = int(h / 2)
w_split_offset = int(w / 2)

pad_w = new_w - w
pad_h = new_h - h

img_array = np.asarray(img)
if h is not new_h:
    t_img = img_array[:h_split_offset, :, :]
    t_img = np.pad(t_img, ((0, pad_h), (0, 0), (0, 0)),
mode=self.PADDING_MODE)
    b_img = img_array[h_split_offset:, :, :]
    img_array = np.concatenate([t_img, b_img], axis = 0)

if w is not new_w:
    l_img = img_array[:, :w_split_offset, :]
    l_img = np.pad(l_img, ((0, 0), (0, pad_w), (0, 0)),
mode=self.PADDING_MODE)
    r_img = img_array[:, w_split_offset:, :]
    img_array = np.concatenate([l_img, r_img], axis = 1)
input_x = img_array
```

[도25]



[도26]

```

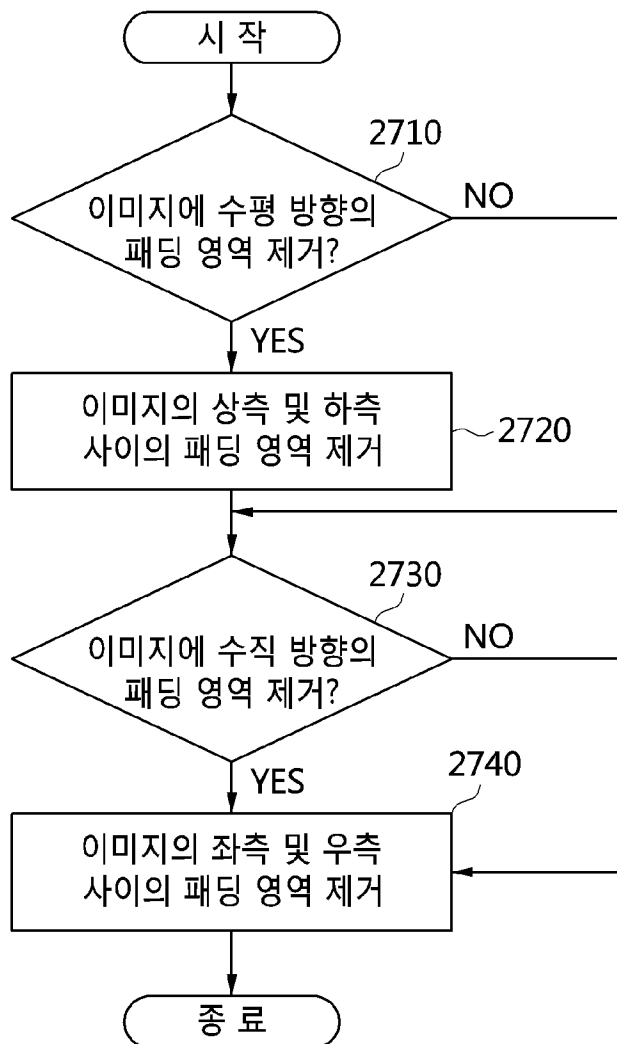
new_w = int(math.ceil(float(w) / float(K)) * K)
new_h = int(math.ceil(float(h) / float(K)) * K)

h_split_offset = int(float(h) / float(2))
w_split_offset = int(float(w) / float(2))

pad_w = new_w - w
pad_h = new_h - h

if h is not new_h:
    recon = np.concatenate([recon[:h_split_offset, :, :],
recon[h_split_offset + pad_h:, :, :]], axis=0)
if w is not new_w:
    recon = np.concatenate([recon[:, :w_split_offset, :], recon[:,
w_split_offset + pad_w:, :]], axis=1)
  
```

[도27]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/007039

## A. CLASSIFICATION OF SUBJECT MATTER

*H04N 1/41(2006.01)i, G06N 20/00(2019.01)i*

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N 1/41; G06N 3/04; H04N 19/117; H04N 19/124; H04N 19/159; H04N 19/169; H04N 19/176; H04N 19/61; G06N 20/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
 Korean utility models and applications for utility models: IPC as above  
 Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 eKOMPASS (KIPO internal) & Keywords: entropy, model, context, gaussian, machine learning, weight, standard deviation

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEE, Jooyoung et al. Context-adaptive Entropy Model for End-to-end Optimized Image Compression. arXiv: 1809.10452v4. 06 May 2019, [Retrieved on 14 August 2020], Retrieved from <URL: <a href="https://arxiv.org/abs/1809.10452">https://arxiv.org/abs/1809.10452</a> >, pages 1-20 See pages 1-2, 6, 13; and figure 3.	1,6,10-12,17
Y		2-5,7,13-16,18
A		8-9,19-20
Y	US 2017-0105005 A1 (QUALCOMM INCORPORATED) 13 April 2017 See paragraph [0130].	2-5,13-16
Y	US 10225607 B1 (NOVATEK MICROELECTRONICS CORP.) 05 March 2019 See claims 1, 5-6.	7,18
A	KR 10-2018-0001428 A (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 04 January 2018 See claims 1, 19.	1-20
A	CN 106713935 A (HANGZHOU DIANZI UNIVERSITY) 24 May 2017 See claim 1.	1-20

 Further documents are listed in the continuation of Box C. See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;" document member of the same patent family


Date of the actual completion of the international search

25 AUGUST 2020 (25.08.2020)

Date of mailing of the international search report

26 AUGUST 2020 (26.08.2020)

Name and mailing address of the ISA/KR



Korean Intellectual Property Office  
 Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu,  
 Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

**PCT/KR2020/007039**

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2017-0105005 A1	13/04/2017	CN 108141583 A	08/06/2018
		CN 108141584 A	08/06/2018
		EP 3360324 A1	15/08/2018
		EP 3360325 A1	15/08/2018
		US 2017-0105004 A1	13/04/2017
		WO 2017-062373 A1	13/04/2017
		WO 2017-062377 A1	13/04/2017
US 10225607 B1	05/03/2019	CN 110087084 A	02/08/2019
KR 10-2018-0001428 A	04/01/2018	CN 109644268 A	16/04/2019
		JP 2019-525544 A	05/09/2019
		US 2019-0230354 A1	25/07/2019
		WO 2017-222140 A1	28/12/2017
CN106713935 A	24/05/2017	CN 106713935 B	11/06/2019

<b>A. 발명이 속하는 기술분류(국제특허분류(IPC))</b> H04N 1/41(2006.01)i, G06N 20/00(2019.01)j		
<b>B. 조사된 분야</b> 조사된 최소문헌(국제특허분류를 기재) H04N 1/41; G06N 3/04; H04N 19/117; H04N 19/124; H04N 19/159; H04N 19/169; H04N 19/176; H04N 19/61; G06N 20/00 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC		
국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 엔트로피(entropy), 모델(model), 문맥(context), 가우시안(gaussian), 기계 학습(machine learning), 가중치(weight), 표준편차(standard deviation)		
<b>C. 관련 문헌</b>		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	JOOYOUNG LEE 등, 'Context-adaptive Entropy Model for End-to-end Optimized Image Compression', arXiv:1809.10452v4, 2019.05.06, [검색일 2020.08.14]. 출처 <URL : https://arxiv.org/abs/1809.10452>, 페이지 1-20 페이지 1-2, 6, 13; 및 도면 3	1, 6, 10-12, 17
Y		2-5, 7, 13-16, 18
A		8-9, 19-20
Y	US 2017-0105005 A1 (QUALCOMM INCORPORATED) 2017.04.13 단락 [0130]	2-5, 13-16
Y	US 10225607 B1 (NOVATEK MICROELECTRONICS CORP.) 2019.03.05 청구항 1, 5-6	7, 18
A	KR 10-2018-0001428 A (한국과학기술원) 2018.01.04 청구항 1, 19	1-20
A	CN 106713935 A (HANGZHOU DIANZI UNIVERSITY) 2017.05.24 청구항 1	1-20
<input type="checkbox"/> 추가 문헌이 C(계속)에 기재되어 있습니다. <input checked="" type="checkbox"/> 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: "A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 "D" 본 국제출원에서 출원인이 인용한 문헌 "E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후 "X"에 공개된 선출원 또는 특허 문헌 "L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 "O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 "P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 "T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 "X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. "Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. "&" 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2020년 08월 25일 (25.08.2020)	국제조사보고서 발송일 2020년 08월 26일 (26.08.2020)	
ISA/KR의 명칭 및 우편주소  대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 양정록 전화번호 +82-42-481-5709	

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2017-0105005 A1	2017/04/13	CN 108141583 A CN 108141584 A EP 3360324 A1 EP 3360325 A1 US 2017-0105004 A1 WO 2017-062373 A1 WO 2017-062377 A1	2018/06/08 2018/06/08 2018/08/15 2018/08/15 2017/04/13 2017/04/13 2017/04/13
US 10225607 B1	2019/03/05	CN 110087084 A	2019/08/02
KR 10-2018-0001428 A	2018/01/04	CN 109644268 A JP 2019-525544 A US 2019-0230354 A1 WO 2017-222140 A1	2019/04/16 2019/09/05 2019/07/25 2017/12/28
CN106713935 A	2017/05/24	CN 106713935 B	2019/06/11