

Deep Learning for Image Super-Resolution: A Survey

인공지능학과 21110690 오윤서

Abstract

이미지 초해상도(Image Super Resolution, SR)는 컴퓨터 비전에서 이미지와 비디오의 해상도를 향상시키는 중요한 이미지 처리 기술이다. 최근 몇 년 동안 Deep Learning 기술을 사용하여 이미지 초해상도의 놀라운 발전을 목격했다. 딥 러닝 접근법을 사용하여 이미지 SR의 최근 발전에 대한 포괄적인 조사를 제공하는 것을 목표로 한다. 일반적으로 SR 기술에 대한 기존 연구를 supervised SR, unsupervised, SR and domain-specific SR의 세 가지 주요 범주로 대략 그룹화할 수 있다. 또한 공개적으로 사용 가능한 benchmark datasets 및 성능 평가 metric과 같은 몇 가지 중요한 문제도 다룬다. 마지막으로, 앞으로 지역사회가 더 다루어야 할 몇 가지 미래 방향과 열린 문제들을 강조함으로써 마무리한다.

1. Introduction

저해상도(LR) 이미지에서 고해상도(HR) 이미지를 복구하는 과정을 말하는 Image SR은 컴퓨터 비전 및 이미지 처리에서 이미지 처리 기술의 중요한 부류이다. 이것은 의료 이미징[1],[2],[3], 감시 및 보안[4],[5]과 같은 다양한 실제 응용 프로그램을 사용한다. 이미지 시각 품질을 향상시키는 것 외에도 다른 컴퓨터 비전 작업[6],[7],[8],[9]을 개선하는 데 도움이 된다. 일반적으로 이 문제는 단일 LR 이미지에 해당하는 여러 개의 HR 이미지가 항상 있기 때문에 매우 어렵고 본질적으로 잘못 제기된다. 여러 문헌에서는 예측 기반 방법[10],[11],[12], 가장자리 기반 방법[13],[14], 통계적 방법[15],[16], 패치 기반 방법[13],[17],[18],[19] 및 희소 표현 방법[20],[21] 등 다양한 고전적 SR 방법이 제안

되었다.

본 논문에서는 딥 러닝을 이용한 Image SR의 최근 발전에 대한 포괄적인 개요를 제시한다. 기존의 SR조사가 많지만, 딥 러닝 기반의 SR 기법에 초점을 맞추고 있다는 점에서 차이가 있지만, 대부분의 초기 연구는 전통적인 SR 알고리즘을 조사하거나 일부 연구는 주로 전체 참조 matric이나 인간의 시각적 인식을 기반으로 정량적 평가를 제공하는 데 초점을 맞추고 있다. 기존의 survey와는 달리, 이 survey는 SR 기법의 최근 발전을 체계적이고 포괄적으로 검토하기 위해 독특한 심층 학습 기반 관점을 취한다.

주요 공헌은 세가지 이다.

1. 딥 러닝 기반의 영상 초해상도 기법에 대해 문제 설정, benchmark datasets, 성능 metric, 딥 러닝을 적용한 SR 메소드 제품군, 도메인별 SR API 등을 포함하여 종합적으로 검토한다.
2. 딥 러닝 기반 SR 기법의 최근 발전을 계층적, 구조적, 체계적으로 개괄하고, 효과적인 SR 솔루션을 위한 각 구성 요소의 장점과 한계를 요약한다.
3. 도전과 열린 문제를 논의하고, 지역 사회에 통찰력 있는 지침을 제공하기 위해 새로운 동향과 미래의 방향을 확인한다.

본 survey는 딥 러닝을 통해 Image SR의 최근 발전에 대한 다양한 측면을 중심으로 다룬다. 섹션 2에서는 문제 정의를 제공하고 주류 datasets 및 metric을 검토한다. 섹션 3에서는 supervised SR의 주요 구성요소를 모듈화 하여 분석한다. 섹션 4에서는 unsupervised SR 방법에 대한 간략한 소개를 제공한다. 섹션 5에서

는 인기있는 domain-specific SR 응용 프로그램을 소개하고 섹션 6에서는 향후 방향과 공개 문제에 대해 설명한다.

2 PROBLEM SETTING AND TERMINOLOGY

2.1 Problem Definitions

2.2 Datasets for Super-Resolution

오늘날 Image SR에는 이미지 양, 품질, 해상도 및 다양성 등이 크게 다른 다양한 datasets가 있다. 이들 중 일부는 LR-HR 이미지 쌍을 제공하는 반면, HR 이미지만을 제공하거나, LR 이미지만 제공하는 데 이 경우는 일반적으로 MATLAB의 기본 설정(즉, bicubic interpolation with anti-aliasing) 기능을 통해 얻어진다. 표 1에서는 SR 커뮤니티에서 일반적으로 사용하는 여러 이미지 데이터 세트를 나열하고 HR 이미지의 양, 평균 해상도, 평균 픽셀 수, 이미지 형식 및 카테고리 키워드를 구체적으로 나타낸다.

이러한 datasets 외에도 ImageNet[51], MS-COCO[52], VOC2012[53], CelebA[54]와 같이 다른 비전 작업에 널리 사용되는 일부 데이터 세트도 SR에 사용된다. 또한, T91과 BSDS300[26],[27],[55],[56], DIV2K와 Flickr2K[31],[57]를 결합하는 등 훈련을 위한 여러 데이터셋을 결합하는 것도 인기가 있다.

2.3 Image Quality Assessment

이미지 품질은 이미지의 시각적 속성을 말하며 시청자의 지각 평가에 중점을 둔다. 일반적으로 IQA(Image Quality Assessment) 방법에는 사람의 인식(즉, 이미지가 얼마나 사실적으로 보이는지)에 기초한 주관적인 방법과 객관적인 계산 방법이 포함된다. 전자(주관적 계산 방법)는 우리의 필요에 더 부합하지만 종종 시간이 많이 걸리고 비용이 많이 들기 때문에 후자(객관적 계산 방법)가 현재 주류를 이루고 있다. 그러나 객관적인 방법은 종종 인간의 시각적 인식을 매우 정확하게 포착할 수 없고 IQA 결과에 큰 차이를 유발할 수 있기 때문에 이러한 방법이 서로 일치할 필요는 없다 [25],[58].

또한 객관적인 IQA 방법은 참조 이미지를 사용하여

평가를 수행하는 전체 참조 방법, 추출된 특징의 비교에 기반한 참조 감소 방법 및 참조 이미지가 없는 참조 없는 방법(즉, 블라인드 IQA)의 세 가지 유형으로 더 나뉜다[58].

2.4 Operating Channels

일반적으로 사용되는 RGB 색 공간 외에도 YCbCr 색 공간도 SR에 널리 사용된다. 이 공간에서 이미지는 Y, Cb, Cr 채널로 표시되며 각각 휘도, 파란색 차이 및 빨간색 차이 chroma 구성 요소를 나타낸다. 현재 어떤 공간에서 초해상도를 수행하거나 평가하기 위한 최선의 방법은 없지만 이전 모델은 YCbCr 공간 [26],[43],[78],[79]의 Y 채널에서 작동하는 것을 선호하지만 최신 모델은 RGB채널[28],[31],[57],[70]에서 작동하는 경향이 있다. 다른 색상 공간이나 채널에서 작동(훈련 또는 평가)하면 평가 결과가 크게 달라질 수 있다. 다른 색상 공간이나 채널에서 작동(훈련 또는 평가)하면 평가 결과가 크게 달라질 수 있다 (최대 4dB)[23].

2.5 Super-Resolution Challenges

이 섹션에서는 Image SR, NTIRE 및 PIRM에 대한 두 가지 가장 일반적인 문제를 간략하게 소개한다.

NTIRE challenge. 이미지 복원 및 향상 (NTIRE) 과제의 새로운 추세[80]는 CVPR과 관련되어 있으며 SR, 노이즈 제거 및 컬러화와 같은 여러 작업을 포함한다. Image SR의 경우 NTIRE challenge는 DIV2K[42] datasets를 기반으로 하며 사실적인 알려지지 않은 저하가 있는 bicubic down scaling 트랙과 blind 트랙으로 구성된다. 이러한 트랙은 저하 및 스케일링 요인이 다르며 이상적인 조건과 실제 불리한 상황에서 SR 연구를 촉진하는 것을 목표로 한다.

PIRM challenge. PIRM (Perceptual Image Restoration and Manipulation) 과제는 ECCV와 관련되어 있으며 여러 작업을 포함한다. NTIRE와 달리 PIRM의 하위 과제 중 하나는 생성 정확도와 지각 품질 간의 절충에 초점을 맞추고 다른 하나는 스마트 폰의 SR에 초점을 맞춘다. 잘 알려진 것처럼[77] 왜곡 대상 모델은 시각적으로 불쾌한 결과를 자주 생성하는 반면, 지각 품질 대상 모델은 정보 충실도에 대해 제대로 수행되지 않는다. 특히, PIRM은 RMSE (root mean squared error)의 임계 값에 따라 인식 왜곡 평면을 세 영역으로 나누었다. 각 영역에서 우승 알고리즘은 NIQE[76]과 Ma[66]

에 의해 평가된 최고의 지각 품질[77]을 달성하는 알고리즘이다. 다른 하위 과제[81]인 스마트 폰의 SR에서 참가자는 제한된 스마트 폰 하드웨어 (CPU, GPU, RAM 등 포함)로 SR을 수행하도록 요청 받고 평가 metric에는 PSNR, MS-SSIM 및 MOS 테스트가 포함된다. 이러한 방식으로 PIRM은 인식 왜곡 트레이드 오프에 대한 고급 연구를 장려하고 스마트 폰에서 가볍고 효율적인 이미지 향상을 유도한다.

3 SUPERVISED SUPER-RESOLUTION

요즘 연구자들은 딥 러닝을 이용한 다양한 초해상도 모델을 제안했다. 이 모델은 supervised SR, 즉 LR 이미지와 해당 HR 이미지로 모두 훈련된 SR에 초점을 맞춘다. 이러한 모델 간의 차이는 매우 크지만 기본적으로 모델 프레임 워크, upsampling 방법, 네트워크 설계 및 학습 전략과 같은 구성 요소 집합의 일부 조합이다. 이러한 관점에서 연구자들은 이러한 구성 요소를 결합하여 특정 목적에 맞는 통합 SR 모델을 구축한다. 이 섹션에서는 각 모델을 개별적으로 도입하는 대신 기본 구성 요소 (그림 1에서 보여주듯이)를 모듈식으로 분석하고 장점과 한계를 요약하는 데 중점을 둔다.

3.1 Super-Resolution Frameworks

Image SR은 잘못된 문제이므로 upsampling (즉, LR 입력에서 HR 출력을 생성하는)을 수행하는 방법이 핵심 문제이다. 기존 모델의 아키텍처는 매우 다양하지만, 채택된 upsampling 작업과 모델 내 위치를 기반으로 하는 네 가지 모델 프레임워크(그림 2에서 보듯이)에 기인할 수 있다.

3.2 Upsampling Methods

모델의 upsampling 위치 외에도 upsampling을 수행하는 방법이 매우 중요하다. 다양한 전통적인 upsampling 방법[20],[21],[88],[89]이 있었지만, 종단 간 upsampling을 배우기 위해 CNN을 활용하는 것이 점차 하나의 트렌드가 되었다. 이 섹션에서는 전통적인 interpolation 기반 알고리즘과 딥 러닝 기반 upsampling 레이어를 소개한다.

3.3 Network Design

오늘날 네트워크 설계는 딥 러닝의 가장 중요한 부

분 중 하나이다. 초해상도 분야에서 연구자들은 최종 네트워크를 구성하기 위해 4 개의 SR 프레임 워크 (섹션 3.1) 위에 모든 종류의 네트워크 설계 전략을 적용한다. 이 섹션에서는 이러한 네트워크를 네트워크 설계를 위한 필수 원칙이나 전략으로 분해하여 소개하고 장점과 한계를 하나씩 분석한다.

3.4 Learning Strategies

3.5 Other Improvements

네트워크 설계 및 학습 전략 외에도 SR 모델을 더욱 향상시키는 다른 기술이 있다.

3.6 State-of-the-Art Super-Resolution Models

최근에는 deep learning 기반의 영상 초해상도 모델이 점점 주목을 받으며 최첨단 성능을 실현하고 있다. 이전 섹션에서는 SR 모델을 모델 프레임 워크 (섹션 3.1), upsampling 방법 (섹션 3.2), 네트워크 설계 (섹션 3.3) 및 학습 전략 (섹션 3.4)을 포함한 특정 구성 요소로 분해하고 이러한 구성 요소를 계층적으로 분석하고 장점과 한계를 식별한다. 사실, 오늘날 대부분의 최첨단 SR 모델은 기본적으로 위에서 요약한 여러 전략의 조합에 기인할 수 있다. 예를 들어, RCAN[70]의 가장 큰 기여는 채널 주의 메커니즘 (섹션 3.3.5)에서 비롯되며 하위 픽셀 upsampling (섹션 3.2.2), 잔여 학습 (섹션 3.3.1), 픽셀 L1 손실 (섹션 3.4.1) 및 자체 앙상블 (섹션 3.5.5)과 같은 다른 전략을 사용한다. 이와 유사한 방식으로, 우리는 표 2에서 보여주는 것처럼 몇 가지 대표적인 모델과 주요 전략을 요약한다.

SR 정확도 외에도 효율성은 또 다른 매우 중요한 측면이며 다양한 전략이 효율성에 어느 정도 영향을 미친다. 따라서 이전 섹션에서는 제시된 전략의 정확성을 분석할 뿐만 아니라 포스트 upsampling (섹션 3.1.2), 재귀 학습 (섹션 3.3.2), 조밀 연결 (섹션 3.3.4), xUnit (섹션 3.3.11)과 같이 효율성에 더 큰 영향을 미치는 전략에 대한 구체적인 영향을 나타낸다. 또한 SR 정확도 (즉, PSNR), 모델 크기 (즉, 매개 변수 수) 및 계산 비용 (즉, Multi-Adds의 수)에 대한 대표적인 SR 모델을 벤치마킹한다(그림 8). 정확도는 4 개의 벤치마크 데이터 세트 (즉, Set5[48], Set14[49], B100[40] 및 Urban100[50])에 대한 PSNR의 평균으로 측정된다. 그리고 모델 크기와 계산 비용은 PyTorch-OpCounter[157]로 계산된다. 여기서 출력 해상도는 720p (즉, 1080-720)이다. 모든 통계는 원본 논문에서

파생되거나 공식 모델에서 계산되며 스케일링 계수는 2이다. 더 나은 보기 및 비교를 위해, 우리는 또한 대화형 온라인 버전을 제공한다.

4 UNSUPERVISED SUPER-RESOLUTION

기존의 초해상도 작업은 주로 감독 학습, 즉 일치된 LR-HR 이미지 쌍으로 학습하는 데 초점을 맞춘다. 그러나 동일한 장면이지만 서로 다른 해상도의 영상을 수집하는 것은 어렵기 때문에 SR 데이터 세트의 LR 영상은 HR 영상에 대해 미리 정의된 열화를 수행하여 얻는 경우가 많다. 따라서 훈련된 SR 모델은 실제로 미리 정의된 열화의 역 과정을 학습한다. 수동 열화 사전은 도입하지 않고 실제 LR-HR 매핑을 배우기 위해 연구자들은 감독되지 않은 SR에 점점 더 많은 관심을 기울인다. 이 경우 훈련을 위해 pairing 되지 않은 LR-HR 이미지만 제공되므로 결과 모델은 실제 시나리오에서 SR 문제에 대처할 가능성이 높다. 다음으로 우리는 딥 러닝을 사용하여 몇 가지 기존의 감독되지 않은 SR 모델을 간략하게 소개할 것이며 더 많은 방법이 아직 탐구되지 않았다.

4.1 Zero-Shot Super-Resolution

단일 이미지 내부의 내부 이미지 통계가 SR에 대한 충분한 정보를 제공한다는 것을 고려할 때, Shocheret et al.[83] 대규모 외부 데이터 세트에 대한 일반 모델을 교육하기보다는 테스트 시간에 이미지 특정 SR 네트워크를 교육하여 감독되지 않은 SR에 대처하기 위한 zero-shot super-resolution(ZSSR)을 제안한다. 특히, 그들은[158]을 사용하여 단일 이미지에서 열화 커널을 추정하고 이 커널을 사용하여 다른 스케일링 factor로 열화를 수행하고 이 이미지에 증강을 수행하여 작은 데이터 세트를 작성한다. 그런 다음 SR을 위한 작은 CNN이 데이터 세트에서 훈련되어 최종 예측에 사용된다.

이러한 방식으로 ZSSR은 모든 이미지 내부의 교차 스케일 내부 반복을 활용하므로 이상적이지 않은 조건(즉, bicubic 분해로 얻은 이미지 및 실제 장면에 가까운 흐림, 노이즈, 압축 artifact와 같은 영향을 받은 이미지)에서 이미지에 대한 이전 접근 방식보다 큰 마진(추정 커널의 경우 1dB, 알려진 커널의 경우 2dB)을 능가한다. 이상적인 조건(즉, bicubic 분해에 의해 얻어진 이미지) 하에서 경쟁적인 결과를 제공한다. 그러

나 테스트 중에 서로 다른 이미지에 대해 서로 다른 네트워크를 훈련해야 하기 때문에 추론 시간이 다른 것보다 훨씬 길다.

4.2 Weakly-Supervised Super-Resolution

미리 정의된 저하를 도입하지 않고 초해상도에 대처하기 위해 연구자들은 약한 감독 학습, 즉 pairing 되지 않은 LR-HR 이미지를 사용하여 SR 모델을 학습하려고 시도한다. 그 중 일부 연구자들은 먼저 HR-to-LR 분해를 학습하여 SR 모델을 교육하기 위한 데이터 세트를 구성하는 반면, 다른 연구자들은 LR-to-HR 및 HR-to-LR 매핑을 동시에 학습하기 위해 사이클인 사이클 네트워크를 설계한다. 다음에는 이 모델들을 자세히 설명하겠다.

Learned Degradation. 미리 정의된 열화는 차선택이기 때문에, 짝이 없는 LR-HR 데이터 세트로부터 열화를 학습하는 것이 실현 가능한 방향이다. Bulat et al.[159]는 먼저 HR-to-LR GAN을 훈련하여 pairing 되지 않은 LR-HR 이미지를 사용하여 열화를 학습한 다음 첫 번째 GAN을 기반으로 수행된 paired LR-HR 이미지를 사용하여 SR 용 LR-to-HR GAN을 훈련시키는 2 단계 프로세스를 제안한다. 특히, HR-to-LR GAN의 경우, HR 이미지를 생성기에 공급하여 LR 출력을 생성하는데, 이는 HR 이미지를 다운 스케일링(평균 pooling)하여 얻은 LR 이미지 뿐만 아니라 실제 LR 이미지의 분포도 일치시켜야 한다. 훈련을 마친 후 생성기는 열화 모델로 사용되어 LR-HR 이미지 쌍을 생성한다. 그런 다음 LR-to-HR GAN의 경우 생성기(즉, SR 모델)는 생성된 LR 이미지를 입력으로 사용하고 해당 HR 이미지 뿐만 아니라 HR 이미지의 분포를 일치시키는 데 필요한 HR 출력을 예측한다.

이 2단계 프로세스를 적용함으로써 제안된 비 감독 모델은 초해상도 실제 LR 이미지의 품질을 효과적으로 향상시키고 이전의 최첨단 작업에 비해 큰 향상을 얻는다.

Cycle-in-Cycle Super-Resolution. 감독되지 않은 초해상도에 대한 또 다른 접근법은 LR 공간과 HR 공간을 두 개의 도메인으로 처리하고 사이클 간 구조를 사용하여 서로 간의 매핑을 학습하는 것이다. 이 경우 교육 목표에는 mapping된 결과를 대상 도메인 분포와 일치시키고 왕복 매핑을 통해 이미지를 복구할 수 있도록 하는 것이 포함된다.

CycleGAN에 의해 동기 부여[138], Yuan et al.[131]은 4 개의 생성기와 2 개의 판별기로 구성된 Cycle-in-cycle SR 네트워크 (CinCGAN)를 제안하며, 잡음이 많은 LR \times clean LR 및 clean LR \times clean HR 매핑을 위해 각각 2 개의 CycleGAN을 구성한다. 특히, 제1 CycleGAN에서, 잡음이 많은 LR 이미지는 생성기로 공급되고, 출력은 실제 깨끗한 LR 이미지의 분포와 일치해야 한다. 그런 다음 다른 발전기에 공급되어 원래 입력을 복구해야 한다. 사이클 일관성, 분배 일관성 및 매핑 유효성을 보장하기 위해 여러 손실 함수 (예 : 적대적 손실, 사이클 일관성 손실, 신원 손실)가 사용된다. 다른 CycleGAN은 매핑 도메인이 다르다는 점을 제외하고는 유사하게 설계되었다.

미리 정의된 열화를 피하기 때문에 감독되지 않은 CinCGAN은 감독된 방법과 비교할 만한 성능을 달성할 뿐만 아니라 매우 가혹한 조건에서도 다양한 경우에 적용할 수 있다. 그러나 SR 문제의 본질과 CinCGAN의 복잡한 아키텍처로 인해 훈련 난이도와 불안정성을 줄이기 위한 몇 가지 고급 전략이 필요하다.

4.3 Deep Image Prior

CNN 구조가 역 문제에 앞서 많은 low-level 이미지 통계를 포착하기에 충분하다는 것을 고려할 때, Ulyanov et al.[160] SR을 수행하기 전에 수작업으로 무작위로 초기화된 CNN을 사용한다. 특히, 임의의 벡터 z 를 입력으로 사용하고 대상 HR 이미지 I_y 를 생성하려고 하는 생성기 네트워크를 정의한다. 목표는 downsampled \hat{I}_y 가 LR 이미지 I_x 와 동일한 \hat{I}_y 를 찾기 위해 네트워크를 훈련시키는 것이다. 네트워크가 무작위로 초기화되고 결코 훈련되지 않기 때문에 유일한 이전은 CNN 구조 자체이다. 이 방법의 성능은 감독된 방법 (2dB)보다 여전히 나쁘지만 전통적인 bicubic up-sampling (1dB)보다 월등히 뛰어나다. 또한 CNN 아키텍처 자체의 합리성을 보여주며 심층 학습 방법론과 CNN 구조 또는 자기 유사성과 같은 수공예 사전 결합하여 SR을 개선하도록 유도한다.

5 DOMAIN-SPECIFIC APPLICATIONS

5.1 Depth Map Super-Resolution

Depth map은 장면에서 시점과 객체 사이의 깊이(즉,

거리)를 기록하고 포즈 추정[161],[162] 및 의미 분할[163],[164]와 같은 많은 작업에서 중요한 역할을 한다. 그러나 경제적 및 생산상의 제약으로 인해 깊이 센서에 의해 생성된 depth map은 종종 저해상도이며 노이즈, 양자화 및 누락된 값과 같은 저하 효과를 겪는다. 따라서 depth map의 공간 해상도를 높이기 위해 초해상도가 도입되었다.

요즘 depth map SR에 대한 가장 보편적인 관행 중 하나는 다른 경제적인 RGB 카메라를 사용하여 LR 깊이 map을 슈퍼 해상도로 안내하기 위해 동일한 장면의 HR 이미지를 얻는 것이다. 구체적으로 Song et al.[165]는 depth map과 RGB 이미지 사이의 깊이 필드 통계 및 로컬 상관 관계를 이용하여 전역 통계 및 로컬 구조를 제한한다. Hui et al.[166] 두 개의 CNN을 사용하여 LR depth map을 upsampling하고 HRRGB 이미지를 다운 샘플링한 다음 RGB 기능을 동일한 해상도의 upsampling depth map에 대한 지침으로 사용한다. 그리고 Haefner et al.[167] 또한 색상 정보를 이용하고 shape-shading기술에 의존하여 SR을 안내한다. 대조적으로, Riegler et al.[168] CNN과 에너지 최소화 모델을 강력한 변형 모델의 형태로 결합하여 다른 참조 이미지 없이 HR depth map을 복구한다.

5.2 Face Image Super-Resolution

얼굴 이미지 초해상도, 일명. 얼굴 환각 (FH)은 종종 다른 얼굴 관련 작업을 도울 수 있다[72],[73],[169]. 일반적인 이미지와 비교하여 얼굴 이미지는 더 많은 얼굴 관련 구조 정보를 가지므로 FH에 얼굴 사전 지식 (예 : 랜드 마크, 파싱 map, ID)을 통합하는 것이 매우 인기 있고 유망한 접근법이다.

가장 간단한 방법 중 하나는 생성된 이미지가 동일한 얼굴 관련 속성을 기본 진리로 갖도록 제한하는 것이다. 구체적으로, CBN[170]은 FH 및 조밀 대응 필드 추정을 교대로 최적화함으로써 얼굴 사전을 이용한다. Super-FAN[30]과 MTUN[171]은 모두 FAN을 도입하여 엔드 투 엔드 multitasking 학습을 통해 얼굴 랜드 마크의 일관성을 보장한다. 그리고 FSRNet[73]은 얼굴 랜드마크 히트map뿐만 아니라 얼굴 파싱 map도 이전의 제약조건으로 사용한다. SICNN[72]는 실제 정체성 회복을 목표로 하며, 초정체성 상실 함수와 도메인 통합 훈련 접근법을 채택하여 공동 훈련을 안정적으로 수행한다.

얼굴 사전을 명시 적으로 사용하는 것 외에도 암묵적

방법도 널리 연구되고 있다. TDN[172]은 자동 공간 변환을 위한 공간 변환기 네트워크[173]를 통합하고, 따라서 얼굴 비 정렬 문제를 해결한다. TDN에 기초하여, TDAE[174]는 encoder-decoder decoder 프레임워크를 채택하는데, 여기서 첫 번째 decoder는 upsample 및 잡음 제거를 학습하고, 인코더는 이를 정렬되고 잡음 없는 LR면으로 다시 투사하며, 마지막 decoder는 hallucinated HR 이미지를 생성한다. 대조적으로, LCGE[175]는 5 개의 얼굴 구성 요소에 대해 SR을 수행하기 위해 구성 요소 별 CNN을 사용하고, HR 얼굴 구성 요소 데이터 세트에서 k-NN 검색을 사용하여 해당 패치를 찾고, 미세한 구성 요소를 합성하고, 최종적으로 FH 결과에 융합한다. 마찬가지로 Yang et al.[176] deblocked 얼굴 이미지를 얼굴 구성 요소와 배경으로 분해하고, 구성 요소 랜드 마크를 사용하여 외부 데이터 세트에서 적절한 HR 예제를 검색하고, 배경에서 일반 SR을 수행하고, 마지막으로 HR 얼굴을 완성하기 위해 융합한다.

또한 연구자들은 다른 관점에서 FH를 개선한다. 인간 주의 이동 메커니즘[112]에 의해 동기 부여된 Attention-FH[113]는 유인 얼굴 패치를 순차적으로 발견하고 로컬 향상을 수행하기 위해 순환 정책 네트워크에 의존하므로 얼굴 이미지의 글로벌 상호 의존성을 완전히 활용한다. URDGN[177]은 적대적 학습을 하는 SRGAN[25]과 유사한 네트워크를 채택한다. 그리고 Xu et al.[63]는 일반 생성기와 클래스 별 판별기로 구성된 다중 클래스 GAN 기반 FH 모델을 제안한다. Lee et al.[178]와 Yu et al.[179]는 조건부 GAN[180]에 기초하여, 지정된 속성을 갖는 FH를 수행하기 위해 추가적인 얼굴 속성 정보를 이용한다.

5.3 Hyperspectral Image Super-Resolution

panchromatic 이미지 (PAN, 즉 3 밴드의 RGB 이미지)와 비교하여 수백 개의 밴드를 포함하는 hyperspectral images (HSIs)는 풍부한 스펙트럼 기능을 제공하고 다양한 비전 작업을 돕는다[181],[182],[183]. 하지만 HD적인 한계로 인해 고품질의 HSI를 수집하는 것은 PAN에 비해 훨씬 어렵고 해상도 또한 떨어진다. 따라서 초해상도가 이 분야에 도입되었으며 연구자들은 HR PAN과 LR HSI를 결합하여 HR HSI를 예측하는 경향이 있다. 그 중 Masi et al.[184]는 SRCNN[22]를 사용하고 성능을 향상시키기 위해 비선형 방사성 지표의 여러 map을 통합한다. Qu et al.[185]는 PAN과 HSI에서

각각 SR을 수행하기 위해 두 개의 인코더-decoder 네트워크를 공동으로 훈련시키고, decoder를 공유하고 각도 유사성 손실 및 재구성 손실과 같은 제약 조건을 적용함으로써 SR 지식을 PAN에서 HSI로 전송한다. 최근에 Fu et al.[186]은 HSI-SR에 대한 카메라 스펙트럼 응답 (CSR) 기능의 효과를 평가하고 최적의 CSR을 자동으로 선택하거나 설계할 수 있는 CSR 최적화 계층을 제안하고 최첨단 기술을 능가한다.

5.4 Real-World Image Super-Resolution

일반적으로, SR 모델들을 트레이닝하기 위한 LR 이미지들은 RGB 이미지들을 수동으로 다운 샘플링(예를 들어, bicubic 다운 샘플링)함으로써 생성된다. 그러나 실제 카메라는 실제로 12bit 또는 14bit RAW 이미지를 캡처하고 카메라 ISP (이미지 신호 프로세서)를 통해 일련의 작업 (예 : demosaicing, 노이즈 제거 및 압축)을 수행하고 최종적으로 8 비트 RGB 이미지를 생성한다. 이 과정을 통해 RGB 이미지는 원본 신호를 많이 잃어버렸고 카메라로 찍은 원본 이미지와 매우 다르다. 따라서 수동으로 downsampled RGB 이미지를 SR에 직접 사용하는 것은 차선책이다.

이 문제를 해결하기 위해 연구자들은 실제 이미지를 SR에 사용하는 방법을 연구한다. 그 중에서도 첸 등 (Chen et al.[187] 이미지 시스템에서 이미지 해상도 (R)와 시야 (V) 사이의 관계를 분석 (즉, R-V 저하) 실제 데이터 세트 City100을 수행하기 위한 데이터 수집 전략을 제안하고 실험적으로 제안된 이미지 합성 모델의 우수성을 입증한다. Zhang et al.[188]은 카메라의 광학 줌을 통해 또 다른 실제 이미지 데이터 세트 SR-RAW (즉, 쌍을 이루는 HR RAW 이미지 및 LR RGB 이미지)를 구축하고, 오정렬 문제를 해결하기 위해 문맥 양자 손실을 제안한다. 대조적으로, Xu et al.[189]는 이미징 프로세스를 시뮬레이션 하여 현실적인 훈련 데이터를 생성하고 RAW 이미지에서 원래 캡처된 광도 정보를 활용하기 위해 이중 CNN을 개발하는 파이프라인을 제안한다. 또한 효과적인 색상 보정 및 다른 센서로의 일반화를 위해 공간적으로 다양한 색상 변환을 학습할 것을 제안한다.

5.5 Video Super-Resolution

비디오 수퍼 해상도의 경우 여러 프레임이 훨씬 더 많은 장면 정보를 제공하며 프레임 내 공간 의존성 뿐만 아니라 프레임 간 시간 의존성 (예 : 동작, 밝기 및

색상 변경)이 있다. 따라서 기존의 연구들은 명시적인 움직임 보상(광학적 흐름 기반, 학습 기반)과 반복적인 방법 등 시공간적 의존성을 더 잘 활용하는 데 초점을 맞추고 있다.

Optical-flow 기반 방법 중 Liao et al.[190] HR 후보자를 생성하고 CNN에 의해 양상불하기 위해 광학 흐름 방법을 사용한다. VSRnet[191]과 CVSRnet[192]은 Druleas 알고리즘[193]에 의한 움직임 보상을 다루며, CNN을 이용하여 연속적인 프레임을 입력으로 받아 HR 프레임을 예측한다. 반면에 Liu et al.[194],[195]는 정류된 광학 흐름 정렬을 수행하고, 다양한 시간 스케일에서 HR 프레임을 생성하고 적응적으로 집계하기 위한 시간 적응적 망을 제안한다.

뿐만 아니라, 다른 사람들은 모션 보상을 직접 배우려고 노력한다. VESPCN[156]은 인접 프레임을 기반으로 모션 보상을 학습하기 위해 훈련 가능한 공간 변환기[173]를 사용하고, 종단 간 예측을 위해 시공간적 ESPCN[84]에 다중 프레임을 입력한다. 그리고 Tao et al.[196] 정확한 LR 이미징 모델로부터 루트를 추출하고 모션 보상과 초해상도를 동시에 달성하기 위해 서브 픽셀과 같은 모듈을 제안하여 정렬된 프레임을 보다 효과적으로 융합한다.

또 다른 추세는 반복적인 방법을 사용하여 명시적인 모션 보상 없이 시공간 종속성을 포착하는 것이다. 구체적으로, BRCN[197],[198]은 양방향 프레임워크를 사용하고, CNN, RNN 및 조건부 CNN을 사용하여 각각 공간적, 시간적 및 공간-시간적 의존성을 모델링한다. 유사하게, STCN[199]은 공간 및 시간 정보를 추출하기 위해 깊은 CNN 및 양방향 LSTM[200]을 사용한다. 그리고 FRVSR[201]은 이전에 추론된 HR 추정치를 사용하여 두 개의 심층 CNN에 의해 후속 HR 프레임을 반복적으로 재구성한다. 최근에 FSTRN[202]은 원래의 대형 필터를 대체하기 위해 두 개의 훨씬 작은 3D convolution 필터를 사용하므로 낮은 계산 비용을 유지하면서 더 깊은 CNN을 통해 성능을 향상시킵니다. RBPN[87]은 반복 encoder-decoder에 의해 공간 및 시간 컨텍스트를 추출하고 이를 back-projection 메커니즘에 기반한 반복 정제 프레임 워크와 결합합니다 (섹션 3.1.4).

또한 FAST[203]은 압축 알고리즘에 의해 추출된 구조 및 픽셀 상관 관계에 대한 간략한 설명을 활용하고 한 프레임에서 인접한 프레임으로 SR 결과를 전송하

며 성능 손실이 거의 없는 최첨단 SR 알고리즘을 훨씬 가속화한다. 그리고 Joet al.[204]는 각 픽셀의 로컬 시공간적 이웃들에 기초하여 동적 upsampling 필터들 및 HR 오차 이미지를 생성하고, 또한 명시적 움직임 보상을 회피한다.

5.6 Other Applications

딥 러닝 기반의 초해상도는 다른 도메인 특정 응용 프로그램에도 적용되며 뛰어난 성능을 보여준다. 특히, Perceptual GAN[205]은 작은 물체의 표현을 초해상도로 처리하여 큰 물체와 유사한 특성을 가지며 탐지를 위해 더 차별적인 작은 물체 탐지 문제를 해결한다. 마찬가지로 FSR-GAN[206]은 픽셀 공간 대신 feature 공간에서 작은 크기의 이미지를 수퍼 분해하여 원시 불량 feature를 매우 차별적인 feature로 변환하므로 이미지 검색에 큰 도움이 된다. 게다가, Jeon et al.[207]은 스테레오 이미지에서 시차를 이용하여 등록 시 서브 픽셀 정확도로 HR 이미지를 재구성한다. Wang et al.[208]은 스테레오 이미지 초해상도 문제를 해결하기 위한 시차주의 모델을 제안한다. Li et al.[209]는 3D 기하학적 정보를 통합하고 3D 객체 텍스처 map을 초분해한다. 그리고 Zhang et al.[210] 하나의 라이트 필드에 있는 뷰 이미지를 그룹으로 분리하고, 모든 그룹에 대한 고유 매핑을 학습하고, 마지막으로 모든 그룹의 오차를 결합하여 고해상도 라이트 필드를 재구성한다. 전반적으로 초고해상도 기술은 모든 종류의 애플리케이션에서 중요한 역할을 할 수 있는데, 특히 우리가 큰 물체를 잘 다룰 수 있지만 작은 물체를 다룰 수 없을 때 그렇다.

6 CONCLUSION

본 논문에서는 딥 러닝을 이용한 이미지 초해상도의 최근 발전에 대한 광범위한 조사를 실시하였다. 우리는 주로 감독 및 감독되지 않은 SR의 개선에 대해 논의했으며 일부 도메인 별 응용 프로그램도 소개했다. 큰 성공에도 불구하고, 아직 해결되지 않은 문제들이 많이 있다. 본 연구가 연구자들에게 이미지 SR에 대한 이해를 높일 수 있을 뿐만 아니라 향후 이 분야의 연구 활동 및 응용 개발에 도움이 되기를 기대한다.

REFERENCES

[1] Zhihao Wang , Jian Chen , and Steven C. H. Hoi,
"Deep Learning for Image Super-Resolution: A Survey,"
IEEE Transactions on Pattern Analysis and Machine
Intelligence, 2021, pp. 3388-3415