가상 시착을 위한 스타일 기반 글로벌 외관 흐름

센허(Sen He), 송이제(Yi-Zhe Song), 타오샹(Tao Xiang)
서리 대학교 시각, 음성 및 신호 처리 센터
iFlyTek-Surrey 인공지능 공동연구센터
{센헤,와이.송,티샹}@surrey.ac.uk

그림 1. 우리의 글로벌 외관 흐름 기반 시도 모델은 Cloth-flow와 같은 기존 로컬 흐름 기반 SOTA 방법에 비해 분명한 이점을 가지고 있습니다. 13] 및 PF-AFN [10], 특히 참조 이미지와 의복 이미지(맨 위 행) 사이에 큰 정렬 오류가 있고 어려운 포즈/가림(아래 행)이 있는 경우에 그렇습니다.

추상적인

이미지 기반 가상 시착은 매장 내 의류를 옷을 입은 사람 이미지에 맞추는 것을 목표로 합니다. 이를 달성하기 위한 핵심 단계는 대상 의류를 인물 이미지의 해당 신체 부위와 공간적으로 정렬하는 의류 워핑입니다. 이전 방법은 일반적으로 로컬 모양 흐름 추정 모델을 채택합니다. 따라서 그들은 본질적으로 어려운 신체 자세/폐색 및 사람과 의복 이미지 사이의 큰 잘못된 정렬에 취약합니다(그림 1 참조).1). 이러한 한계를 극복하기 위해 본 연구에서는 새로운 전역 출현 흐름 추정 모델을 제안합니다. 처음으로 StyleGAN 기반 아키텍처가 모양 흐름 추정에 채택되었습니다. 이를 통해 앞서 언급한 문제에 대처하기 위해 전체 이미지 컨텍스트를 인코딩하는 글로벌 스타일 벡터를 활용할 수 있습니다. StyleGAN 흐름 생성기가 로컬 의류 변형에 더 많은 주의를 기울이도록 안내하기 위해 흐름 개선 모듈이 도입되어 로컬 컨텍스트를 추가합니다. 인기 있는 가상 Tryon 벤치마크의 실험 결과는 우리의 방법이 새로운 최첨단 성능을 달성했음을 보여줍니다. 이는 참조 이미지가 전신이어서 의류 이미지와 크게 정렬되지 않는 '야생' 애플리케이션 시나리오에서 특히 효과적입니다(그림 1).1맨 위). 코드는 다음에서 확인할 수 있습니다.https: //github.com/SenHe/Flow-Style-VTON.

1. 소개

최근 팬데믹으로 인한 폐쇄로 인해 오프라인 매장 소매에서 전자상거래로의 전환이 가속화되었습니다. 2020년 전 세계 소매 전자상거래 매출은 4조 2800억 달러에 이르렀고 2022년에는 전자 소매 수익이 5조 4000억 달러로 성장할 것으로 예상됩니다. 라인 쇼퍼는 의류 품목을 입어볼 수 있는 탈의실입니다. 온라인 소매업체의 반품 비용을 줄이고 쇼핑객에게 온라인과 동일한 오프라인 경험을 제공하기 위해 최근 이미지 기반 가상 체험판(VTON)이 집중적으로 연구되고 있습니다.9,10,13,14,19,24,38,39,42,43].

VTON 모델은 매장 내 의류를 인물 이미지에 맞추는 것을 목표로 합니다. VTON 모델의 주요 목적은 매장 내 의류를 인물 이미지의 해당 신체 부위에 정렬하는 것입니다. 이는 매장 내 의류가 일반적으로 사람 이미지와 공간적으로 정렬되지 않기 때문입니다(그림 1 참조).1). 공간 정렬 없이 고급 디테일 보존 이미지를 이미지 변환 모델에 직접 적용 [18,30] 사람 이미지와 의복 이미지의 질감을 융합하면 생성된 입어보기 이미지, 특히 가려지고 잘못 정렬된 영역에서 비현실적인 효과가 발생합니다.

이전 방법은 의류 워핑을 통해 이 정렬 문제를 해결합니다. 즉, 먼저 매장 내 의류를 워핑한 다음 사람 이미지와 연결하고 최종 시착 이미지 생성을 위해 이미지 대 이미지 변환 모델에 입력합니다. 그들 중 다수는 [9,14,19,38,42, 43] 박판스플라인(TPS) 채택 [7] Warping 방법을 기반으로 사람과 의복 이미지에서 추출된 특징 간의 상관 관계를 활용합니다. 그러나 이전 작품에서 분석한 바와 같이 [5,13,42], TPS는 예를 들어 의류의 서로 다른 영역에서 서로 다른 변형이 필요한 경우 복잡한 뒤틀림을 처리하는 데 한계가 있습니다. 그 결과, 최근 SOTA 방식 [10,13] 조밀한 출현 흐름 추정 [45] 옷을 휘게 합니다. 여기에는 의류를 해당 신체 부위에 정렬하는 데 필요한 변형을 나타내는 조밀한 외관 유동장을 예측하기 위한 네트워크 훈련이 포함됩니다.

그러나 기존의 외관 흐름 추정 방법은 전역적 맥락이 부족하여 정확한 의류 워핑에 한계가 있습니다. 보다 구체적으로, 기존의 모든 방법은 로컬 기능의 대응(예: 로컬 기능 연결 또는 상관 관계)을 기반으로 합니다.[1], 광학 흐름 추정을 위해 개발됨 [6,17]. 외관 흐름을 추정하기 위해 그들은 인물 이미지와 매장 내 의류의 해당 영역이 특징 추출기의 동일한 로컬 수용 필드에 위치한다는 비현실적인 가정을 합니다. 의복과 해당 신체 부위 사이의 정렬이 크게 어긋난 경우(그림 1).1위) 현재의 외관 흐름 기반 방법은 급격히 저하되고 만족스럽지 못한 결과를 낳게 됩니다. 글로벌 컨텍스트가 부족하면 기존 흐름 기반 VTON 방법이 어려운 포즈/폐색에 취약해집니다(그림 1).1하단) 해당 지역을 넘어 서신을 검색해야 하는 경우. 이는 '야생' 방법의 사용을 심각하게 제한하며, 이에 따라 사용자는 여러 의류 품목(예: 상의, 하의 및 신발)을 입어보기 위해 개인 이미지로서 자신의 전신 사진을 가질 수 있습니다.

이러한 한계를 극복하기 위해 본 연구에서는 새로운 전역 출현 흐름 추정 모델을 제안합니다. 특히, 처음으로 StyleGAN [21,22] 조밀한 출현 흐름 추정을 위한 아키텍처. 이는 기존 방식과 근본적으로 다르다.6,10,13,17] U-Net을 사용하는 [30] 지역 공간적 맥락을 보존하기 위한 아키텍처. 전체 참조 및 의류 이미지에서 추출된 글로벌 스타일 벡터를 사용하면 모델이 글로벌 컨텍스트를 쉽게 캡처할 수 있습니다. 그러나 이는 또한 중요한 질문을 제기합니다. 지역 정렬에 중요한 지역 공간적 맥락을 포착할 수 있습니까? 결국 단일 스타일 벡터는 로컬 공간 컨텍스트를 잃어버린 것 같습니다. 이 질문에 답하기 위해 먼저 StyleGAN이 성공적으로 수행되었다는 점에 주목합니다.

다양한 스타일 벡터가 다양한 시점에서 동일한 얼굴을 생성할 수 있는 로컬 얼굴 이미지 조작 작업에 적용됩니다.34] 그리고 다양한 모양 [15,28]. 이는 전역 스타일 벡터에 로컬 공간 컨텍스트가 인코딩되어 있음을 나타냅니다. 그러나 우리는 또한 바닐라 StyleGAN 아키텍처가 [21,22]는 U-Net에 비해 큰 정렬 오류와 어려운 포즈/폐색에 대해 훨씬 더 강력하지만 로컬 변형 모델링에서는 약합니다. 따라서 우리는 기존 StyleGAN 생성기에 로컬 흐름 개선 모듈을 도입하여 두 세계의 장점을 모두 갖췄습니다.

구체적으로 StyleGAN 기반 워핑 모듈(*여*그림에서2)는 전역 스타일 벡터, 의복 특징 및 사람 특징을 입력으로 사용하는 누적된 워핑 블록으로 구성됩니다. 글로벌 스타일 벡터는 글로벌 컨텍스트 모델링을 위해 사람 이미지와 매장 내 의류의 최저 해상도 특징 맵에서 계산됩니다. 생성기의 각 워핑 블록에서 전역 스타일 벡터는 모양 흐름을 추정하기 위해 해당 의류 특징 맵을 가져오는 특징 채널을 변조하는 데 사용됩니다. 흐름 추정기가 세밀한 국부적 모양 흐름(예: 그림의 팔과 손 영역)을 모델링할 수 있도록 합니다.5, 스타일 기반 외관 흐름 추정 부분 위에 각 워핑 블록에 개선 레이어를 도입합니다. 이 정제 레이어는 먼저 의류 특징 맵을 왜곡한 후 동일한 해상도로 사람 특징 맵과 연결한 다음 로컬 세부 모양 흐름을 예측하는 데 사용됩니다.

기여(1) 가상 시착 시 의상을 변형시키는 새로운 스타일 기반 외관 흐름 방법을 제안합니다. 이러한 전역 흐름 추정 접근 방식을 통해 VTON 모델은 사람과 의복 이미지 사이의 큰 잘못된 정렬에 대해 훨씬 강력해졌습니다. 이는 우리의 방법을 자연스러운 포즈를 가진 전신 인물 이미지가 사용되는 '야생' 애플리케이션에 더 적용 가능하게 만듭니다(그림 1 참조).1). (2) 우리는 우리의 방법을 검증하기 위해 광범위한 실험을 수행하여 이 방법이 기존의 최첨단 대안보다 우수하다는 것을 분명히 보여줍니다.

2. 관련 업무

이미지 기반 가상 체험이미지 기반(2D) VTON은 파서 기반 방법과 파서 프리 방법으로 분류할 수 있습니다. 주요 차이점은 상용 인간 파서인지 여부입니다.[2]추론 단계에서 필요합니다.

파서 기반 방법은 왜곡 매개변수 추정을 위해 입력 사람 이미지에서 의류 영역을 마스크하기 위해 사람 분할 맵을 적용합니다. 마스크를 쓴 사람 이미지는 뒤틀린 의복과 연결된 다음 대상 시착 이미지 생성을 위한 생성기에 공급됩니다. 대부분의 방법 [9,13,14,38,42,43] 사전 훈련된 인간 파서를 적용합니다.11]

[1]텐서 상관 방법이 [6,10,17] 글로벌 수용 분야에 도달할 가능성이 있습니다. 그러나 계산은 입력 크기에 따라 2차적으로 증가합니다. 다루기 쉽도록 하기 위해 실제 구현은 여전히 제한된 지역 환경을 기반으로 합니다.

[2]때로는 미리 훈련된 포즈 [삼] 및 조밀한 포즈 [12] 탐지 모델은 파서 기반 모델에도 사용됩니다.

사람 이미지를 사전 정의된 여러 의미 영역(예: 머리, 상의, 바지)으로 구문 분석합니다. 더 나은 시착 이미지 생성을 위해 [42]는 또한 대상 의류와 일치하도록 분할 맵을 변환합니다. 변형된 파싱 결과는 뒤틀린 의복 및 마스크를 쓴 사람 이미지와 함께 최종 시착 이미지 생성에 사용됩니다. 파서에 대한 의존도는 이러한 방법을 인간의 나쁜 파싱 결과에 민감하게 만듭니다. 10,19] 이는 필연적으로 부정확한 뒤틀림 및 시착 결과를 초래합니다.

대조적으로, 파서가 없는 메소드 [10,19], 추론 단계에서는 사람 이미지와 의복 이미지만 입력으로 사용합니다. 이는 잘못된 구문 분석 결과로 인한 부정적인 영향을 제거하기 위해 특별히 설계되었습니다. 이러한 방법은 일반적으로 먼저 파서 기반 교사 모델을 훈련한 다음 파서가 없는 학생 모델을 정제합니다. [19]는 쌍을 이루는 삼중항을 사용하여 의류 워핑 모듈과 시착 생성 네트워크를 증류하는 파이프라인을 제안했습니다. [10] 더욱 개선되었습니다 [19] 더 나은 증류를 위해 주기 일관성을 도입했습니다. 우리의 방법은 파서가 없는 방법이기도 합니다.

그러나 우리의 방법은 의류 워핑 부분의 디자인에 중점을 두고 있으며 여기서는 새로운 글로벌 외관 흐름 기반 의류 워핑 모듈을 제안합니다.

3D 가상 체험이미지 기반 VTON에 비해 3D VTON은 더 나은 시험 경험을 제공하지만(예: 임의의 보기 및 포즈로 볼 수 있음) 더 어렵습니다. 대부분의 3D VTON 작동 [2,27] 3D 파라메트릭 인체 모델에 의존 [25] 훈련을 위해 스캔된 3D 데이터세트가 필요합니다. 대규모 3D 데이터 세트를 수집하는 것은 비용이 많이 들고 힘들기 때문에 3D VTON 모델의 확장성에 제약이 됩니다. 이러한 문제를 극복하기 위해 최근 [44] 비모수적 이중 인간 깊이 모델 적용 [8] 단안부터 3D VTON까지. 그러나 기존 3D VTON은 여전히 2D 방법에 비해 열악한 텍스처 디테일을 생성합니다.

이미지 조작을 위한 StyleGAN스타일GAN [21,22]는 이미지 조작에 대한 연구에 혁명을 일으켰습니다.28,33,41] 최근에. 고도로 풀린 잠재 공간을 학습하는 데 적합하기 때문에 이미지 조작 작업에 성공적으로 적용되는 경우가 많습니다. 최근의 노력은 감독되지 않은 잠재 의미론 발견에 집중되어 왔습니다.4, 34,37]. [24] 가상 체험을 위해 포즈 조건을 갖춘 StyleGAN을 적용했습니다. 그러나 해당 모델은 의류 세부 사항을 보존할 수 없으며 추론 중에 속도가 느립니다.

우리 의류 워핑 네트워크의 디자인은 이미지 조작, 특히 모양 변형의 뛰어난 성능을 갖춘 StyleGAN에서 영감을 얻었습니다.28,34]. 뒤틀린 의복을 생성하기 위해 스타일 변조를 사용하는 대신, 스타일 변조를 사용하여 암시적 모양 흐름을 예측한 다음 샘플링을 통해 의복을 뒤틀는 데 사용됩니다. 이 디자인은 [에 비해 의복의 디테일을 보존하는 데 훨씬 더 적합합니다.24].

외관 흐름VTON의 맥락에서 출현 흐름은 [13]. 그 이후로 이득을 얻었습니다 최근 최신 VTON 모델에 채용되어 더욱 주목받고 있습니다.5, 10]. 기본적으로 외관 흐름은 의류 워핑을 위한 샘플링 그리드로 사용되므로 정보 손실이 없고 세부 보존이 우수합니다. VTON 외에도 외관 흐름은 다른 작업에서도 인기가 있습니다. [45] 새로운 관점 합성을 위해 적용했습니다. [1,29] 또한 사람 포즈 전송을 위한 특징 맵을 왜곡하기 위해 모양 흐름의 아이디어를 적용했습니다. 기존의 모든 출현 흐름 추정 방법과 달리, 우리의 방법은 스타일 변조를 통해 전역 스타일 벡터를 적용하여 출현 흐름을 추정합니다. 따라서 우리의 방법은 큰 정렬 불량에 대처하는 능력이 본질적으로 우수합니다.

3. 방법론

3.1. 문제 정의

사람 이미지가 주어지면 ($\mathit{피} \in$아르 자형$^{삼 \times 시간 \times 여}$)매장 내 의류 이미지($\mathit{g} \in$아르 자형$^{삼 \times 시간 \times 여}$),가상 시착의 목표는 시착 이미지를 생성하는 것입니다($\mathit{티} \in$아르 자형$^{삼 \times 시간 \times 여}$)옷이 어디에 있는지g해당 부분에 맞는$\mathit{피}$. 또한, 생성된$\mathit{티}$, 두 세부정보 모두g비의류 지역$\mathit{피}$보존되어야 합니다. 즉, 같은 사람이$\mathit{피}$변경되지 않은 상태로 나타나야 합니다.$\mathit{티}$지금 입고 있는 것 외에는g.

부정확한 인간 구문 분석의 부정적인 영향을 제거하기 위해 제안된 모델($\mathit{에프}$그림에서2)은 파서가 없는 모델로 설계되었습니다. 기존의 파서가 없는 모델이 채택한 전략을 따릅니다. 10,19], 먼저 파서 기반 모델을 사전 훈련합니다($\mathit{에프}_{PB}$). 그런 다음 파서가 없는 최종 모델을 훈련하는 데 도움이 되는 지식 증류의 교사로 사용됩니다.$\mathit{에프}$. 둘 다$\mathit{에프}$그리고$\mathit{에프}_{PB}$세 부분으로 구성, 즉 두 가지 기능 추출기($\mathit{이자형}_{PB}$/$\mathit{이자형}_{\mathit{피}}$~에$\mathit{에프}_{PB}$그리고$\mathit{이자형}_{\mathit{피}}$,$\mathit{이자형}_{g}$~에$\mathit{에프}$), 뒤틀림 모듈($\mathit{여}_{PB}$~에$\mathit{에프}_{PB}$그리고$\mathit{여}$~에$\mathit{에프}$) 및 생성기($G_{PB}$ ~에$\mathit{에프}_{PB}$그리고G~에$\mathit{에프}$). 각각에 대해서는 다음 섹션에서 자세히 설명합니다.

3.2. 파서 기반 모델 사전 학습

기존 파서가 없는 모델의 표준에 따라 [10,19], 파서 기반 모델$\mathit{에프}_{PB}$처음으로 훈련받습니다. 제안된 파서 프리 모델의 후속 학습에서는 두 가지 방식으로 사용됩니다.$\mathit{에프}$: (a) 사람 이미지를 생성합니다($\mathit{피}$)에 의해 사용됩니다$\mathit{에프}$입력으로 (b) 교육을 감독합니다.$\mathit{에프}$지식 증류를 통해.

구체적으로,$\mathit{에프}_{PB}$의미론적 표현(분할 맵)을 입력으로 사용합니다.삼, 키포인트 포즈 및 밀집 포즈) 실제 인물 이미지($\mathit{피}_{gt} \in$아르 자형$^{삼 \times 시간 \times 여}$)훈련 세트와 짝이 없는 의복($\mathit{g}_{유엔} \in$아르 자형$^{삼 \times 시간 \times 여}$).의 출력 $\mathit{에프}_{PB}$이미지입니다$\mathit{피}$원래 사람이 입고 있던 곳 $\mathit{g}_{유엔}$.$\mathit{피}$에 대한 입력 역할을 할 것입니다.$\mathit{에프}$훈련 중. 이 디자인은 [10], 우리가 지금

삼분할 맵의 의류 영역이 배경 영역으로 반전됩니다.

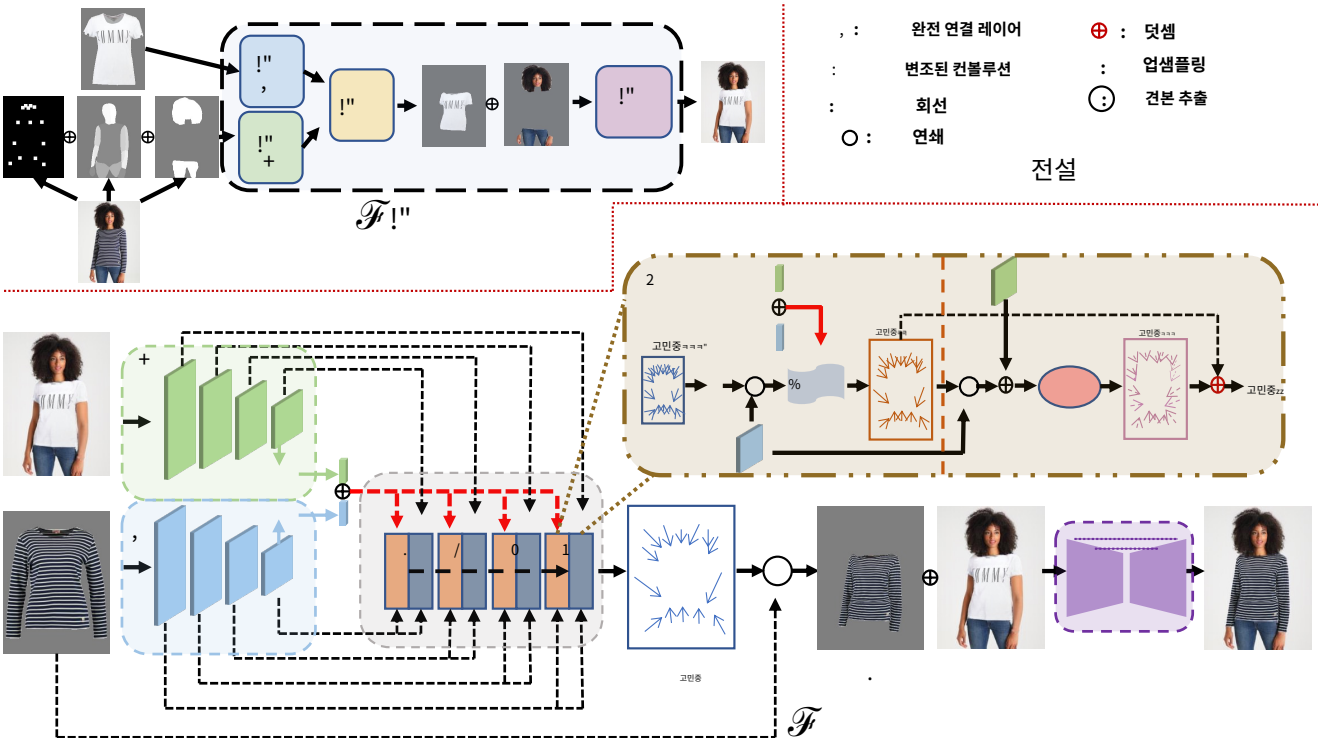

그림 2. 우리 프레임워크의 개략도. 사전 훈련된 파서 기반 모델*에프PB*파서 프리 모델의 입력으로 출력 이미지를 생성합니다.*에프*. 두 가지 특징 추출기는*에프*인물 이미지와 의상 이미지의 특징을 각각 추출합니다. 사람 이미지와 의류 이미지의 가장 낮은 해상도의 특징 맵에서 스타일 벡터를 추출합니다. 워핑 모듈은 사람 이미지와 의복 이미지에서 스타일 벡터와 특징 맵을 가져와서 모양 흐름 맵을 출력합니다. 그런 다음 모양 흐름을 사용하여 의복을 휘게 합니다. 마지막으로, 뒤틀린 의복은 사람 이미지와 연결되어 생성기에 입력되어 대상 시착 이미지를 생성합니다. 참고하세요*에프PB*훈련 중에만 사용됩니다.

페어링된 인물 이미지가 있습니다.*피gt*및 의류 이미지*g*~에*피gt*파서가 없는 모델을 훈련하기 위해*에프*, 그건:

$$\text{에프} = \underset{\text{에프}}{\text{인수 분}} rrt - p_{gt}rr, \quad (1)$$

어디*티*=*에프*(*피, 지*)에서 생성된 시착 이미지입니다.*에프*. 참고하세요*에프PB*훈련 중에만 사용됩니다.*에프*.

3.3. 특징 추출

두 개의 컨벌루션 인코더를 적용합니다(*이자형피*그리고*이자형g*)의 특징을 추출하기 위해*피*그리고*g*. 둘 다*이자형피*그리고*이자형g*누적된 잔여 블록으로 구성된 동일한 아키텍처를 공유합니다. 에서 추출된 특징*이자형피*그리고*이자형g*다음과 같이 표현될 수 있다 $\{피_나\}_N$그리고$\{g_나\}_N$(*N*=4그림에서2단순화를 위해), 여기서 $피_나 \in$아르 자형$^{씨나×시간나×승나}$그리고$g_나 \in$아르 자형$^{씨나×시간나×승나}$는 해당 잔차 블록에서 추출된 특징 맵입니다.*이자형피*그리고*이자형g*, 각각. 추출된 특징 맵은 다음에 사용됩니다.*여*출현 흐름을 예측합니다.

3.4. 스타일 기반 외관 흐름 예측

제안된 모델의 주요 신규 구성 요소는 스타일 기반 전역 외관 흐름 추정 모듈입니다. 이전 방법과 다른점

~예 표준 PP귀 an

지역 특징 대응을 기반으로 [10,13], 원래 광학 흐름 추정에서 제안된 [6,17], 우리의 방법은 전역 스타일 벡터를 기반으로 먼저 스타일 변조를 통해 대략적인 모양 흐름을 추정한 다음 사전을 개선합니다.
로컬 기능 대응을 기반으로 거친 모양 흐름을 나타냅니다.

F에 예시된 바와 같이나 g.2, 워핑 모듈(*여*) 구성 N개의 중첩된 워핑 블록($\{W_나\}_N$ 1),각 블록은 다음과 같습니다. 스타일 기반 모양 흐름 예측 레이어(주황색 사각형)와 로컬 대응 기반 모양 흐름 구체화 레이어(파란색 사각형)로 구성됩니다. 구체적으로 우리는
먼저 전역 스타일 벡터를 추출합니다(*에스*∈아르 자형^{씨}) 그로부터 출력된 기능을 사용하여*N*일(최종) 블록*이자형피*그리고*이자형g*, 다음과 같이 표시됨 $피_N$그리고g_N,처럼:

$$에스 = [에프_피(피_N), 에프_g(g_N)], \quad (2)$$

어디*에프피*그리고*에프g*야? 울리 공동 연결된 레이어와 [· 아 ,]드-N연결을 메모합니다. ~ 안에 삼위일체 추출된 글로바 난 스타일 벡터*에스*4그 사람의 전반적인 정보를 담고 있으며 위치, 구조 등 의복. 스타일 기반 이미지 조작과 유사 [15,28, 33,34], 우리는 전 세계적으로

4인투혁극적으로,*에스*=*에프피*($피_N$)모양 흐름을 생성하기에 충분합니다. 경험적 하지만 우리는 으로 발견한*에스*= [*에프피*($피_N$), *에프g*(g_N)]더 나은 결과를 얻습니다.

스타일 벡터 $에스$뒤틀림에 필요한 변형을 포착합니다. g~ 안으로 $피$. 따라서 이는 모양 흐름 필드를 추정하기 위해 StyleGAN 스타일 생성기의 스타일 변조에 사용됩니다.

보다 구체적으로 각 블록의 스타일 기반 출현 흐름 예측 레이어에서는 $여_{나}$, 우리는 거친 흐름을 예측하기 위해 스타일 변조를 적용합니다.

$$에프_{ci}=전환율_{층}(에스(g_{N+1-나는},유(에프_{나-1})),초), \quad (삼)$$

어디 $전환율_{층}$변조된 컨벌루션을 나타냅니다.21],$에스(\cdot,\cdot)$ 샘플링 연산자입니다.$유$업샘플링 연산자이고, 에프$_{나-1}$∈아르 자형$_{2\times 시간_{나-1}\times 승_{나-1}}$는 마지막 워핑 블록에서 예측된 흐름입니다. 참고로 첫 번째 블록은 $여_1$~에 $여$가장 낮은 해상도의 의류 특징 맵과 스타일 벡터만 사용합니다. 즉,에프$_{c1}$=$전환율_{층}(g_N,초)$. 방정식에서 알 수 있듯이삼, 예측에프$_{ci}$의류 기능 맵과 전역 스타일 벡터에 따라 달라집니다. 따라서 이는 전역적인 수용 영역을 가지며 의복과 인물 이미지 사이의 큰 불일치에 대처할 수 있습니다. 그러나 스타일 벡터로는 $에스$는 전역 표현이므로 절충점으로 국지적인 세밀한 모양 흐름을 정확하게 추정하는 능력이 제한되어 있습니다(그림 1 참조).5). 따라서 거친 흐름은 국부적인 개선이 필요합니다.

정제하다에프$_{ci}$, 각 블록에 로컬 대응 기반 외관 흐름 개선 레이어를 도입합니다.$여_{나}$. 국소적인 세밀한 모양 흐름을 추정하는 것을 목표로 합니다.

$$에프_{리}=전환율([에스(g_{N+1-나는},에프_{ci}),피_{N+1-나는}]), \quad (4)$$

어디에프$_{리}$는 예측된 정제 흐름이고,$전환율$컨볼루션을 나타냅니다. 기본적으로 정제 레이어는 로컬 대응, 즉 동일한 수용 필드에서 뒤틀린 사람 특징과 의복 특징 간의 대응을 통해 정제 흐름을 추정합니다. 참고로 뒤틀린 후에는에프$_{ci}$, 우리는 다음의 해당 지역/특징을 가정할 수 있습니다.$g_{N+1-나는}$그리고 $피_{N+1-나는}$이제 동일한 수용 필드에 위치합니다. 따라서 이전 작품에서 사용했던 지역 대응을 적용할 수 있다.10,13] 국부적인 세밀한 외관 흐름을 예측합니다.

마지막으로, 각 워핑 블록의 출력으로 거친 흐름과 로컬 세밀한 모양 흐름을 함께 추가합니다.

$$에프_{나}=에프_{ci}+에프_{리}. \quad (5)$$

예상되는 출현 흐름에프$_N$마지막 블록부터 $여$의복을 휘게 하는 데 사용됩니다.

$$g=에스(g,에프_N). \quad (6)$$

그리고 뒤틀린 옷g그런 다음 사람 이미지와 연결되어 대상 시착 이미지 생성을 위한 생성기에 입력됩니다.

$$에프=G([\hat{g},피]). \quad (7)$$

발전기G사이에 건너뛰기 연결이 있는 인코더-디코더 아키텍처가 있습니다. 우리는 [의 디자인을 따릅니다.18, 46] 텍스처 디테일 보존에 효과적인 것으로 입증되었습니다.

3.5. 학습 목표

모델을 훈련시키기 위해 먼저 지각 손실을 적용합니다.20] 의 출력 사이$에프$그리고 실측 인물 이미지 $피_{gt}$:

$$엘_{피}=\sum_{나} V_{나}(E)-\phi_{나}(피_{gt})rr, \quad (8)$$

어디$\phi_{나}$은$나_{일}$사전 훈련된 VGG 네트워크의 블록 [35].

워핑 모델의 훈련을 감독하기 위해$여$, 뒤틀린 의복에 손실을 적용합니다.

$$엘_g=\hat{g}-중_g\cdot 피_{gt}rr, \quad (9)$$

어디$중_g$의복 마스크입니다$피_{gt}$상용 인간 분석 모델을 통해 예측되었습니다.

이전 모양 흐름 방법의 표준에 따라 [10,13], 우리는 또한 각 블록의 예측 흐름에 대해 평활도 정규화를 적용합니다.$여$:

$$엘_{아르\ 자형}=\sum_{나} rr\nabla 에프_{나}rr, \quad (10)$$

어디$rr\nabla$에프$_{나}rr$일반화된 Charbonnier 손실 함수입니다.36].

파서 기반 사람 인코더에 대한 입력(분할 맵, 키포인트 포즈 및 밀집 포즈)($이자형_{PB}$)는 파서프리 모델보다 더 많은 의미 정보를 포함합니다.$에프$(사람 이미지), 사람 인코더의 학습을 안내하기 위해 증류 손실을 적용합니다.$이자형_{피}$~에 $에프$:

$$엘_{디}=\sum_{나} rr피_{PB}-피_{나}rr, \quad (11)$$

어디$피_{PB}$와 출력 특징 맵은 다음과 같습니다.$나_{일}$사람 인코더 차단$이자형_{PB}$사전 훈련된 파서 기반 모델 $에프_{PB}$.

전반적인 학습 목표는 다음과 같습니다.

$$엘=\lambda_{피}엘_{피}+\lambda_g엘_g+\lambda_{아르\ 자형}엘_{아르\ 자형}+\lambda_{디}엘_{디}, \quad (12)$$

어디$\lambda_{피},\lambda_g,\lambda_{아르\ 자형}$그리고$\lambda_{디}$네 가지 목표의 균형을 맞추기 위한 하이퍼파라미터를 나타냅니다.

4. 실험

데이터세트VITON 데이터 세트에서 모델을 실험합니다.5[14]. 이전 VTON 작업에서 가장 많이 사용된 데이터셋입니다. VITON에는 다음이 포함된 훈련 세트가 포함되어 있습니다.14 ,221이미지 쌍6그리고 테스트 데이터 세트2,032

5데이터 세트의 사용은 [에서 작성자에 의해 허용되었습니다.14].

6각 쌍은 사람의 이미지와 사람의 옷 이미지를 의미합니다.

한 쌍. 사람과 의복 이미지 모두 해당 해상도입니다. 256 × 192.

또한 무작위로 배치된 사람 이미지에 대한 모델의 견고성을 평가하기 위해 증강된 VITON으로 표시되는 테스트 데이터 세트를 생성합니다(그림 1의 예 참조).4) 원본 데이터 세트의 의류 이미지와 더 큰 정렬 오류가 있습니다. VITON의 대부분의 테스트 사람 이미지는 사람 이미지와 의복이 잘 사전 정렬되도록 잘 배치되어 있으므로(예: 사람 이미지와 의복 이미지의 대부분 해당 영역이 대략 동일한 수용 필드에 위치함) 적합하지 않습니다. 이번 평가를 위해. 구체적으로 증강된 VITON 데이터 세트는 VITON의 테스트 대상 이미지를 이동 및 확대/축소를 통해 무작위로 확대하여 생성됩니다. 특히, 이미지에서 사람의 위치를 이동하여 VITON에서 테스트하는 사람 이미지 1/3을 무작위로 확대하고, 이미지에서 사람을 확대/축소하여 VITON에서 또 다른 1/3 테스트 이미지를 무작위로 확대하고 또 다른 1/3 테스트를 유지합니다. 변함없는 이미지. 이 데이터 세트를 평가할 때 비교된 모든 모델은 사람 이미지 확대를 통해 훈련됩니다.

구현 세부정보우리 모델은 PyTorch에서 구현되었습니다. 단일 Nvidia RTX 2080-Ti GPU를 사용하여 모델을 훈련합니다. 배치 크기를 4로 설정하고 100개의 epoch로 모델을 학습합니다. Adam 옵티마이저를 사용하여 모델을 훈련합니다.23]. 초기 학습률은 다음과 같이 설정됩니다.5 *전자*-4이는 50 epoch 후에 선형적으로 붕괴됩니다. 각 잔여 블록 *이자형*피 그리고 *이자형*g공간 차원을 줄이기 위해 풀링 레이어가 뒤따릅니다. 우리는 설정했다 *N*=5그리고 *씨*=256구현 중. 우리는 이 작업이 승인되면 코드를 공개할 것입니다.

평가 지표 및 기준선우리는 자동 및 수동으로 모델을 평가합니다. 자동 평가에서는 VTON의 표준에 따라 구조 유사성(SSIM)을 사용하여 모델 성능을 평가합니다.40] 및 FID(Fréchet Inception Distance) [16]. 에 따르면 [10,31], 개시점수(IS) [32]는 VTON 이미지를 평가하는 데 적합하지 않으므로 평가에 채택하지 않습니다. 수동(주관적) 평가에서는 Amazon Mechanical Turk(AMT)에 대한 지각 연구를 실행하여 다양한 모델에서 생성된 시착 이미지의 품질을 비교합니다. 입력된 사람 이미지, 의류 이미지, 두 모델에서 생성된 시착 이미지가 주어지면 AMT 작업자에게 어떤 시착 이미지가 더 좋은지 투표하도록 요청했습니다. 각 AMT 작업자에게는 두 모델을 비교하기 위해 무작위로 100개의 이미지가 할당되었습니다. 전 모델 비교 평가에는 AMT 작업자 15명이 참여하였습니다.

우리는 우리의 방법을 다른 파서 기반 방법과 비교합니다. VTON [14], CP-VTON [38], 천류 [13], CP-VTON++ [26], ACGPN [42], DC톤 [9] 및 ZFlow [5]. 또한 SOTA 파서가 없는 방법인 PF-AFN과도 비교합니다.10].

행동 양식	워핑	파서	쌈 ↑	버팀대 ↓
브이톤 [14]	TPS	와이	0.74	55.71
CP-VTON [38]	TPS	와이	0.72	24시 45분
CP-VTON++ [26]	TPS	와이	0.75	21.04
천흐름 [13]	AF	와이	0.84	14.43
ACGPN [42]	TPS	와이	0.84	16.64
디시턴 [9]	TPS	와이	0.83	14.82
PF-AFN [10]	AF	N	0.89	10.09
지플로우 [5]	AF	와이	0.88	15.17
옷감 흐름·[13]	AF	N	0.89	10.73
우리 것	AF	N	0.91	8.89

표 1. VITON에 대한 다양한 모델의 정량적 결과. 뒤틀림은 다양한 모델에서 사용되는 뒤틀림 방법을 나타냅니다. 파서는 추론 중에 모델에 인간 파서가 사용되는지 여부를 나타냅니다. TPS: 박판 스플라인. AF: 등장 흐름.
⋆: 파서 없는 훈련 패러다임으로 재훈련되었습니다.

주요 결과VITON 테스트 데이터 세트의 정량적 결과는 표에 나와 있습니다.1. 우리 모델이 새로운 최첨단 성능을 달성한 것을 볼 수 있습니다. 중요한 것은 이전 SOTA 방법 PF-AFN으로 달성한 이미 낮은 FID 점수(10.09)를 고려할 때 우리 방법은 다음과 같이 이를 더욱 줄일 수 있다는 것입니다.11.9%.한편, 표에서 다음과 같은 관찰을 할 수 있습니다.1. (1) 외관 흐름 기반 워핑 방법은 일반적으로 TPS 기반 워핑 방법보다 성능이 좋습니다. (2) 훈련 시간이 더 많이 걸리지만 파서 프리 방법은 파서 기반 방법보다 훨씬 좋습니다. 제안된 새로운 전역 출현 흐름 추정 방법의 이점을 활용하는 우리 모델은 이전 SOTA 파서 프리 방법(PF-AFN)보다 성능이 뛰어납니다.10] 및 Clothflow [13]) 모든 평가 지표에 적용됩니다. 인간 평가 결과는 표에 나와 있습니다.2. 결과는 표의 결과와 일치합니다.1. 우리 모델은 모든 비교 모델보다 더 나은 성능을 발휘합니다.10%우대율. 다양한 모델의 질적 결과가 그림 1에 나와 있습니다.삼. 전반적으로 우리의 방법은 더 나은 시험 이미지를 생성합니다. 예를 들어, 두 번째와 세 번째 행의 하드 포즈와 폐색.

증강 테스트 데이터 세트의 정량적 결과는 표에 나와 있습니다.삼. 볼 수 있듯이 우리 모델은 증강된 VITON 테스트 데이터세트에서 가장 잘 수행된다는 것을 알 수 있습니다. 중요한 것은 다른 모든 모델의 성능이 급격히 떨어진다는 것입니다. 그리고 우리 모델은 원래 VITON 테스트 데이터 세트와 비교하여 여전히 성능(SSIM 점수)을 유지할 수 있습니다. 질적 예는 그림 1에 나와 있습니다.4. 우리 모델만이 정렬 오류가 크기 때문에 일관되고(예: 의류의 왼쪽 소매) 고품질의 시착 이미지를 생성할 수 있습니다.

절제 연구본 실험에서는 출현 흐름 추정 블록의 설계를 검증합니다(*여나*). 특히, 우리는 먼저 글로벌 스타일 변조(SM) 기반 외관 흐름 추정만으로 방법을 실험합니다.

그림 3. 다양한 모델의 정성적 결과(CP-VTON++ [26], ACGPN [42], PF-AFN [9] 및 우리의 것)을 VITON 테스트 데이터 세트에 적용했습니다.

비교 방법	우대율
CP-VTON++ [26]	12.7% / 87.3%
ACGPN [42]	20.2% / 79.8%
옷감 흐름·[13]	38.5% / 61.5%
AF-PFN [10]	43.2% / 56.8%

표 2. 인간 평가에서 다른 모델과 우리 모델(다른 모델/우리 모델)을 비교한 선호도 비율.

행동 양식	쌈 ↑	버팀대	↓▽쌈/▽버팀대
ACGPN	0.81	20.75	0.003/4.11
옷감 흐름·[13]	0.86	13.05	0.003/2.96
AF-PFN [10]	0.87	12.19	0.002/2.10
우리 것	0.91	9.91	0/1.02

표 3. 증강된 VI-TON에 대한 다양한 모델의 정량적 결과 및 상대적 성능 저하(▽쌈/▽버팀대) 표준 VITON 테스트 데이터세트와 비교.

은, 단지 사용하는 것입니다에프ci방정식에서삼각여나. 그런 다음 정제 흐름(RF) 추정만으로 방법을 실험합니다.에프리방정식에서4각여나. 마지막으로, 먼저 스타일 변조를 통해 전체적으로 모양 흐름을 추정한 다음 로컬 대응을 통해 로컬로 모양 흐름을 개선하는 결합 방법(SM + RF)을 실험합니다. 정량적 결과

표에 나와 있습니다.4. 제안된 글로벌 스타일 변조(SM) 기반 출현 흐름 방법은 로컬 대응 기반 방법보다 성능이 뛰어납니다. 결합하면 성능이 더욱 향상됩니다. 그림에 도시된 바와 같이.5, 로컬 개선 없이 우리의 방법(글로벌 스타일 모듈-

그림 4. 무작위로 배치된 사람 이미지에 대한 다양한 VTON 모델의 견고성을 보여줍니다. 첫 번째 행은 원본 인물 이미지를 입력으로 사용합니다. 두 번째 행은 수직으로 이동된 사람 이미지를 입력으로 사용합니다. ACGPN [42], 천류 [13], PF-AFN [10].

경우에 한함), 소매 부분 등 국소적인 세밀한 외관 흐름을 정확하게 예측할 수 없어 만족스럽지 못한 시착 이미지가 생성되는 경우가 있습니다. 그러나 지역 대응 기반의 출현 흐름 추정만 사용합니다.에프$_{리}$~에$_{여나}$, 해당 영역이 동일한 수용 필드에 위치하지 않는 경우 이 방법은 문제가 발생합니다. 그림에 도시된 바와 같이.6,에프$_{리}$입력된 인물 이미지와 의상 이미지 사이에 큰 어긋남이 있을 경우 외모 흐름을 정확하게 예측할 수 없습니다. 한 번 에프$_{ci}$오정렬을 줄이기 위해 처음 사용되었으므로 우리 모델은 문제를 성공적으로 극복할 수 있습니다.

행동 양식	쌈 ↑	버팀대 ↓
RF	0.89	10.73
에스엠	0.89	9.84
SM + RF	0.91	8.89

표 4. 다양한 외관 흐름 추정 방법을 사용한 VTON 테스트 데이터 세트의 결과$_{여나}$. RF: 지역 대응 기반 흐름 추정. SM: 스타일 변조 기반 흐름 추정.

5. 결론

본 논문에서는 의류를 워핑하기 위한 스타일 기반의 전역적 외관 흐름 추정 방법을 제안했습니다.
미덕알 시착. 스타일 변조를 통한 우리의 방법은 먼저 외관 흐티마름을 전체적으로 평가한 다음
국지적으로 외관 흐름. 우리의 방법은 VITON 벤치마크에서 최고 수준의 성능을 달성했으며 사람과 의복 이미지 사이의 큰 정렬 오류뿐만 아니라 어려운 포즈/가림에 대해서도 더욱 강력합니다. 우리는 우리 방법의 우수성을 보여주기 위해 광범위한 실험을 수행하고 아키텍처 설계를 검증했습니다.

그림 5. 결과 비교에프$_{ci}$에 사용$_{여나}$그리고에프$_{ci}$+에프$_{리}$에 사용$_{여나}$.

그림 6. 결과 비교에프$_{리}$에 사용$_{여나}$그리고에프$_{ci}$+에프$_{리}$에 사용$_{여나}$입력된 인물 이미지와 의상 이미지의 어긋남이 큰 경우.

참고자료

[1] Badour AlBahar, Jingwan Lu, Jimei Yang, Zhixin Shu, Eli Shechtman 및 Jia-Bin Huang. 스타일이 있는 포즈: 조건부 스타일건을 사용하여 디테일을 유지하는 포즈 기반 이미지 합성. ~ 안에 *시그라프 아시아*, 2021.삼

[2] Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt 및 Gerard Pons-Moll. 다중 의류 네트: 이미지를 통해 3D 인물에게 옷을 입히는 방법을 학습합니다. ~ 안에 *ICCV*, 2019.삼

[3] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh. 부품 선호도 필드를 사용한 실시간 다중 사람 2D 포즈 추정. ~ 안에 *CVPR*, 2017.2

[4] 안톤 체렙코프(Anton Cherepkov), 안드레이 보이노프(Andrey Voynov), 아르템 바벤코(Artem Babenko). 의미론적 이미지 편집을 위한 gan 매개변수 공간 탐색. ~ 안에 *CVPR*, 2021.삼

[5] Ayush Chopra, Rishabh Jain, Mayur Hemani 및 Balaji Krishnamurthy. Zflow: 3D 사전을 사용한 게이트형 외관 흐름 기반 가상 시험입니다. ~ 안에 *ICCV*, 2021.2,삼,6

[6] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers 및 Thomas Brox. Flownet: 컨벌루션 네트워크를 사용하여 광학 흐름을 학습합니다. ~ 안에 *CVPR*, 2015.2,4

[7] 장 뒤송. Sobolev 공간에서 회전 불변 준규범을 최소화하는 스플라인입니다. ~ 안에 *여러 변수의 함수에 대한 구성 이론*, 85~100페이지. 스프링거, 1977.2

[8] Valentin Gabeur, Jean-Sébastien Franco, Xavier Martin, Cordelia Schmid 및 Gregory Rogez. 인간 성형: 단일 이미지로부터 비모수적 3D 인간 형태 추정. ~ 안에 *ICCV*, 2019.삼

[9] Chongjian Ge, Yibing Song, Yuying Ge, Han Yang, Wei Liu 및 Ping Luo. 매우 현실적인 가상 시착을 위한 풀린 주기 일관성. ~ 안에 *CVPR*, 2021.1,2,6,7

[10] Yuying Ge, Yibing Song, Ruimao Zhang, Chongjian Ge, Wei Liu 및 Ping Luo. 외관 흐름을 정제하여 파서 없이 가상으로 시험해 볼 수 있습니다. ~ 안에 *CVPR*, 2021.1,2,삼,4,5,6,7,8

[11] Ke Gong, Xiaodan Liang, Dongyu Zhang, Xiaohui Shen 및 Liang Lin. 사람을 들여다보세요: 자기주도 구조의 민감한 학습과 인간 분석의 새로운 벤치마크입니다. ~ 안에 *CVPR*, 2017.2

[12] Rıza Alp Güler, Natalia Neverova 및 Iasonas Kokkinos. Densepose: 야생에서 조밀한 인간 자세 추정. ~ 안에 *CVPR*, 2018.2

[13] Xintong Han, Xiaojun Hu, Weilin Huang 및 Matthew R Scott. Clothflow: 옷을 입은 사람 생성을 위한 흐름 기반 모델입니다. ~ 안에 *ICCV*, 2019.1,2,삼,4,5,6,7,8

[14] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu 및 Larry S Davis. Viton: 이미지 기반 가상 시착 네트워크. ~ 안에 *CVPR*, 2018.1,2,5,6

[15] Sen He, Wentong Liao, Michael Ying Yang, Yi-Zhe Song, Bodo Rosenhahn 및 Tao Xiang. 풀린 수명 얼굴 합성. ~ 안에 *ICCV*, 2021.2,4

[16] 마틴 호이젤, 휴버트 램자우어, 토마스 운터티너, 베른하르트 네슬러, 제프 호흐라이터. 두 가지 시간 규모 업데이트 규칙에 의해 훈련된 Gans는 로컬 내쉬 균형으로 수렴됩니다. ~ 안에 *NeurIPS*, 2017.6

[17] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy 및 Thomas Brox. Flownet 2.0: 심층 네트워크를 통한 광학 흐름 추정의 진화. ~ 안에 *CVPR*, 2017.2,4

[18] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou 및 Alexei A Efros. 조건부 적대 네트워크를 사용한 이미지 간 변환. ~ 안에 *CVPR*, 2017.1,5

[19] Thibaut Issenhuth, Jérémie Mary, Clément Calauzenes. 마스킹할 필요가 없는 것은 마스킹하지 마십시오. 즉, 파서가 필요 없는 가상 체험입니다. ~ 안에 *ECCV*, 2020.1,2,삼

[20] 저스틴 존슨, 알렉상드르 알라히, 리페이페이. 실시간 스타일 전송 및 초해상도에 대한 지각 손실. ~ 안에 *ECCV*, 2016.5

[21] Tero Karras, Samuli Laine 및 Timo Aila. 생성적 적대 네트워크를 위한 스타일 기반 생성기 아키텍처입니다. ~ 안에 *CVPR*, 2019.2,삼,5

[22] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen 및 Timo Aila. 스타일건의 이미지 품질을 분석하고 개선합니다. ~ 안에 *CVPR*, 2020.2,삼

[23] Diederik P Kingma와 지미 바. Adam: 확률론적 최적화를 위한 방법입니다. ~ 안에 *ICLR*, 2015.6

[24] Kathleen M Lewis, Srivatsan Varadharajan 및 Ira Kemelmacher-Shlizerman. Tryongan: 레이어 보간을 통한 신체 인식 시도. *옷*, 40(4):1–10, 2021.1,삼

[25] 매튜 로퍼, 노린 마흐무드, 하비에르 로메로, 제라드 폰스-몰, 마이클 J 블랙. Smpl: 스킨이 적용된 다중 사용자 선형 모델입니다. *옷*, 34(6):1–16, 2015.삼

[26] Matiur Rahman Minar, Thai Thanh Tuan, 안희준, Paul Rosin, Yu-Kun Lai. Cp-vton+: 옷의 모양과 질감을 그대로 유지한 이미지 기반 가상 시착. ~ 안에 *CVPRW*, 2020.6,7

[27] Aymen Mir, Thiemo Alldieck 및 Gerard Pons-Moll. 옷 이미지의 질감을 3D 인간에게 전달하는 방법을 배웁니다. ~ 안에 *CVPR*, 2020.삼

[28] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman 및 Ira Kemelmacher-Shlizerman. 수명 연령 변환 합성. ~ 안에 *ECCV*, 739~755페이지, 2020.2,삼, 4

[29] Yurui Ren, Xiaoming Yu, Junming Chen, Thomas H Li 및 Ge Li. 인물 이미지 생성을 위한 심층 이미지 공간 변환. ~ 안에 *CVPR*, 2020.삼

[30] Olaf Ronneberger, Philipp Fischer 및 Thomas Brox. U-net: 생체의학 이미지 분할을 위한 컨벌루션 네트워크. ~ 안에 *미카이*, 2015.1,2

[31] Mihaela Rosca, Balaji Lakshminarayanan, David Warde-Farley 및 Shakir Mohamed. 자동 인코딩 생성적 적대 신경망을 위한 변형 접근법. *arXiv 사전 인쇄본 arXiv:1706.04987*, 2017.6

[32] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford 및 Xi Chen. 간 훈련 기술이 향상되었습니다. ~ 안에 *NeurIPS*, 2016.6

[33] Yujun Shen, Ceyuan Yang, Xiaoou Tang 및 Bolei Zhou. Interfacegan: gan이 학습한 얽힌 얼굴 표현을 해석합니다. *E/파미*, 2020.삼,4

[34] Yujun Shen과 Bolei Zhou. Gans의 잠재 의미론에 대한 폐쇄형 인수분해. ~ 안에 *CVPR*, 2021.2,삼,4

[35] Karen Simonyan과 Andrew Zisserman. 대규모 이미지 인식을 위한 매우 깊은 컨벌루션 네트워크. ~ 안에 *ICLR*, 2015.5

[36] Deqing Sun, Stefan Roth 및 Michael J Black. 광학 흐름 추정의 현재 사례와 그 뒤에 있는 원리에 대한 정량적 분석. *IJCV*, 106(2):115–137, 2014.5

[37] Christos Tzelepis, Georgios Tzimiropoulos 및 Ioannis Patras. Warpedganspace: gan 잠재 공간에서 비선형 rbf 경로 찾기. ~ 안에 *ICCV*, 2021.삼

[38] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin 및 Meng Yang. 특성을 보존하는 이미지 기반 가상 시착 네트워크를 지향합니다. ~ 안에 *ECCV*, 2018.1,2,6

[39] Jiahang Wang, Tong Sha, Wei Zhang, Zhoujun Li 및 Tao Mei. 마지막 디테일까지: 세밀한 디테일을 갖춘 가상 시착. ~ 안에 *ACMMM*, 2020.1

[40] Zhou Wang, Alan C Bovik, Hamid R Sheikh 및 Eero P Simoncelli. 이미지 품질 평가: 오류 가시성부터 구조적 유사성까지. *팁*, 13(4):600–612, 2004.6

[41] Ceyuan Yang, Yujun Shen 및 Bolei Zhou. 의미론적 계층 구조는 장면 합성을 위한 심층적인 생성 표현에서 나타납니다. *IJCV*, 129(5):1451–1466, 2021.삼

[42] 한양(Han Yang), 루이마오 장(Ruimao Zhang), 샤오바오 궈(Xiaobao Guo), 웨이 리우(Wei Liu), 왕멍 주오(Wangmeng Zuo), 핑 루오(Ping Luo). 이미지 콘텐츠를 적응적으로 생성 및 보존하여 사진처럼 사실적인 가상 체험을 지향합니다. ~ 안에 *CVPR*, 2020.1,2,삼,6,7,8

[43] Yu Ruiyun, Xiaoqi Wang, Xiaohui Xie. Vtnfp: 신체 및 의복 특징을 보존하는 이미지 기반 가상 시착 네트워크입니다. ~ 안에 *ICCV*, 2019.1,2

[44] Fuwei Zhao, Zhenyu Xie, Michael Kampffmeyer, Haoye Dong, Songfang Han, Tianxiang Zheng, Tao Zhang 및 Xiaodan Liang. M3d-vton: 단안-3D 가상 시도 네트워크입니다. ~ 안에 *ICCV*, 2021.삼

[45] Tinghui Zhou, Shubham Tulsiani, Weilun Sun, Jitendra Malik 및 Alexei A Efros. 외관 흐름별로 합성을 봅니다. ~ 안에 *ECCV*, 2016.2,삼

[46] 주준얀, 박태성, 필립 이솔라, 알렉세이 A 에프로스. 주기 일관성이 있는 적대 네트워크를 사용하여 짝을 이루지 않은 이미지 간 변환. ~ 안에 *ICCV*, 2017.5