**Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks**

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros

**Abstract**

Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available. We present an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping G : X → Y such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss. Because this mapping is highly under-constrained, we couple it with an inverse mapping F : Y → X and introduce a cycle consistency loss to enforce F(G(X)) ≈ X (and vice versa). Qualitative results are presented on several tasks where paired training data does not exist, including collection style transfer, object transfiguration, season transfer, photo enhancement, etc. Quantitative comparisons against several prior methods demonstrate the superiority of our approach.

이미지 대 이미지 번역은 정렬된 이미지 쌍의 훈련 세트를 사용하여 입력 이미지와 출력 이미지 간의 매핑을 학습하는 것이 목표인 비전 및 그래픽 문제 클래스입니다. 그러나 많은 작업의 경우 쌍을 이루는 훈련 데이터를 사용할 수 없습니다. 짝을 이루는 예가 없을 때 원본 도메인 X에서 대상 도메인 Y로 이미지를 번역하는 방법을 학습하는 방법을 제시합니다. 우리의 목표는 G(X)의 이미지 분포가 adversarial loss를 사용하여 분포 Y와 구별할 수 없도록 매핑 G : X → Y를 배우는 것입니다. 이 매핑은 제약 조건이 매우 낮기 때문에 역 매핑 F : Y → X와 결합하고 F(G(X)) ≈ X(또는 그 반대)를 적용하기 위해 주기 일관성 손실을 도입합니다. 수집 스타일 전송, 개체 변형, 계절 전송, 사진 향상 등을 포함하여 쌍을 이루는 교육 데이터가 존재하지 않는 여러 작업에 대한 정성적 결과가 표시됩니다. 여러 이전 방법에 대한 정량적 비교는 우리 접근 방식의 우수성을 보여줍니다.

**1. Introduction**

텍스트, 스크린샷, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 1: Given any two unordered image collections X and Y, our algorithm learns to automatically “translate” an image from one into the other and vice versa: (left) Monet paintings and landscape photos from Flickr; (center) zebras and horses from ImageNet; (right) summer and winter Yosemite photos from Flickr. Example application (bottom): using a collection of paintings of famous artists, our method learns to render natural photographs into the respective styles.

그림 1: 정렬되지 않은 두 개의 이미지 컬렉션 X와 Y가 주어지면 우리 알고리즘은 이미지를 자동으로 "변환"하는 방법을 학습합니다. (중앙) ImageNet의 얼룩말과 말; (오른쪽) Flickr의 여름과 겨울 요세미티 사진. 예제 응용 프로그램(하단): 유명 예술가의 그림 모음을 사용하여 우리의 방법은 자연스러운 사진을 각각의 스타일로 렌더링하는 방법을 배웁니다.

What did Claude Monet see as he placed his easel by the bank of the Seine near Argenteuil on a lovely spring day in 1873 (Figure 1, top-left)? A color photograph, had it been invented, may have documented a crisp blue sky and a glassy river reflecting it. Monet conveyed his impression of this same scene through wispy brush strokes and a bright palette.

Claude Monet은 1873년 어느 아름다운 봄날 Argenteuil 근처의 센 강둑에 이젤을 놓았을 때 무엇을 보았습니까(그림 1, 왼쪽 위)? 컬러 사진이 발명되었다면 선명한 푸른 하늘과 그것을 반사하는 유리 같은 강을 기록했을 것입니다. 모네는 이 장면에 대한 자신의 인상을 희미한 붓놀림과 밝은 팔레트를 통해 전달했습니다.

What if Monet had happened upon the little harbor in Cassis on a cool summer evening (Figure 1, bottom-left)? A brief stroll through a gallery of Monet paintings makes it possible to imagine how he would have rendered the scene: perhaps in pastel shades, with abrupt dabs of paint, and a somewhat flattened dynamic range.

모네가 시원한 여름 저녁에 카시스의 작은 항구에서 일어났다면(그림 1, 왼쪽 하단)? 모네의 그림 갤러리를 잠시 둘러보면 그가 그 장면을 어떻게 렌더링했을지 상상할 수 있습니다. 아마도 파스텔 색조와 갑작스런 페인트 칠, 그리고 다소 평평한 다이내믹 레인지일 것입니다.

We can imagine all this despite never having seen a side by side example of a Monet painting next to a photo of the scene he painted. Instead, we have knowledge of the set of Monet paintings and of the set of landscape photographs. We can reason about the stylistic differences between these two sets, and thereby imagine what a scene might look like if we were to “translate” it from one set into the other

모네가 그린 장면의 사진 옆에 나란히 있는 모네 그림의 예를 본 적이 없음에도 불구하고 우리는 이 모든 것을 상상할 수 있습니다. 대신, 우리는 모네의 그림 세트와 풍경 사진 세트에 대한 지식을 가지고 있습니다. 우리는 이 두 세트 사이의 문체 차이에 대해 추론할 수 있으며, 따라서 한 세트에서 다른 세트로 "번역"한다면 장면이 어떻게 보일지 상상할 수 있습니다.

텍스트, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 2: Paired training data (left) consists of training examples , where the correspondence between and exists [22]. We instead consider unpaired training data (right), consisting of a source set and a target set , with no information provided as to which matches which .

그림 2: 쌍을 이루는 훈련 데이터(왼쪽)는 훈련 예제 으로 구성되며, 여기서 와 사이의 대응 관계는 존재합니다[22]. 대신 소스 세트 및 타겟 세트 , 어떤 가 어떤 와 일치하는지에 대한 정보는 제공되지 않습니다.

In this paper, we present a method that can learn to do the same: capturing special characteristics of one image collection and figuring out how these characteristics could be translated into the other image collection, all in the absence of any paired training examples.

이 논문에서는 동일한 작업을 수행할 수 있는 방법을 제시합니다. 하나의 이미지 컬렉션의 특수 특성을 캡처하고 이러한 특성을 다른 이미지 컬렉션으로 변환할 수 있는 방법을 알아내는 것입니다.

This problem can be more broadly described as imageto-image translation [22], converting an image from one representation of a given scene, x, to another, y, e.g., grayscale to color, image to semantic labels, edge-map to photograph. Years of research in computer vision, image processing, computational photography, and graphics have produced powerful translation systems in the supervised setting, where example image pairs are available (Figure 2, left), e.g., [11, 19, 22, 23, 28, 33, 45, 56, 58, 62]. However, obtaining paired training data can be difficult and expensive. For example, only a couple of datasets exist for tasks like semantic segmentation (e.g., [4]), and they are relatively small. Obtaining input-output pairs for graphics tasks like artistic stylization can be even more difficult since the desired output is highly complex, typically requiring artistic authoring. For many tasks, like object transfiguration (e.g., zebra↔horse, Figure 1 top-middle), the desired output is not even well-defined.

이 문제는 이미지를 이미지에서 이미지로의 변환[22]으로 더 광범위하게 설명할 수 있습니다. 즉, 주어진 장면 x의 한 표현에서 다른 표현인 y로 이미지를 변환하는 것입니다. . 컴퓨터 비전, 이미지 처리, 컴퓨터 사진 및 그래픽에 대한 수년간의 연구를 통해 감독 환경에서 강력한 번역 시스템이 생성되었으며, 여기서 예시 이미지 쌍 을 사용할 수 있습니다(그림 2, 왼쪽), 예를 들어 [11, 19, 22, 23, 28, 33, 45, 56, 58, 62]. 그러나 짝을 이루는 훈련 데이터를 얻는 것은 어렵고 비용이 많이 들 수 있습니다. 예를 들어, 의미론적 세분화(예: [4])와 같은 작업을 위한 데이터 세트는 몇 개만 존재하며 상대적으로 작습니다. 예술적 양식화와 같은 그래픽 작업을 위한 입력-출력 쌍을 얻는 것은 원하는 출력이 일반적으로 예술적 저작을 필요로 하는 매우 복잡하기 때문에 훨씬 더 어려울 수 있습니다. 개체 변형(예: 얼룩말↔말, 그림 1 상단-중간)과 같은 많은 작업의 경우 원하는 출력도 잘 정의되지 않습니다.

We therefore seek an algorithm that can learn to translate between domains without paired input-output examples (Figure 2, right). We assume there is some underlying relationship between the domains – for example, that they are two different renderings of the same underlying scene – and seek to learn that relationship. Although we lack supervision in the form of paired examples, we can exploit supervision at the level of sets: we are given one set of images in domain X and a different set in domain Y . We may train a mapping G : X → Y such that the output , is indistinguishable from images y ∈ Y by an adversary trained to classify apart from y. In theory, this objective can induce an output distribution over that matches the empirical distribution pdata(y) (in general, this requires G to be stochastic) [16]. The optimal G thereby translates the domain X to a domain distributed identically to Y . However, such a translation does not guarantee that an individual input x and output y are paired up in a meaningful way – there are infinitely many mappings G that will induce the same distribution over . Moreover, in practice, we have found it difficult to optimize the adversarial objective in isolation: standard procedures often lead to the wellknown problem of mode collapse, where all input images map to the same output image and the optimization fails to make progress [15].

따라서 우리는 짝을 이루는 입력-출력 예제 없이 도메인 간 번역을 학습할 수 있는 알고리즘을 찾습니다(그림 2, 오른쪽). 우리는 도메인 사이에 기본 관계가 있다고 가정하고(예: 동일한 기본 장면의 두 가지 다른 렌더링이라고 가정하고 해당 관계를 배우려고 합니다.) 쌍을 이루는 예의 형태로 감독이 부족하지만 집합 수준에서 감독을 활용할 수 있습니다. 도메인 X에 한 세트의 이미지가 주어지고 도메인 Y에 다른 세트가 제공됩니다. 출력 가 를 y와 별도로 분류하도록 훈련된 적에 의해 이미지 y ∈ Y와 구별할 수 없도록 매핑 G : X → Y를 훈련할 수 있습니다. 이론적으로 이 목적은 경험적 분포 pdata(y)와 일치하는 에 대한 출력 분포를 유도할 수 있습니다(일반적으로 이를 위해서는 G가 확률적이어야 함) [16]. 따라서 최적의 G는 도메인 X를 Y와 동일하게 분포된 도메인 Y로 변환합니다. 그러나 이러한 변환은 개별 입력 x와 출력 y가 의미 있는 방식으로 쌍을 이루는 것을 보장하지 않습니다. 에 대해 동일한 분포를 유도하는 매핑 G가 무한히 많습니다. 게다가, 실제로 우리는 적대적 목표를 단독으로 최적화하는 것이 어렵다는 것을 발견했습니다. 표준 절차는 종종 모든 입력 이미지가 동일한 출력 이미지에 매핑되고 최적화가 진행되지 않는 잘 알려진 모드 붕괴 문제로 이어집니다[15].

These issues call for adding more structure to our objective. Therefore, we exploit the property that translation should be “cycle consistent”, in the sense that if we translate, e.g., a sentence from English to French, and then translate it back from French to English, we should arrive back at the original sentence [3]. Mathematically, if we have a translator G : X → Y and another translator F : Y → X, then G and F should be inverses of each other, and both mappings should be bijections. We apply this structural assumption by training both the mapping G and F simultaneously, and adding a cycle consistency loss [64] that encourages F(G(x)) ≈ x and G(F(y)) ≈ y. Combining this loss with adversarial losses on domains X and Y yields our full objective for unpaired image-to-image translation.

이러한 문제는 우리의 목표에 더 많은 구조를 추가해야 합니다. 따라서 우리는 번역이 "주기 일관성"이 있어야 한다는 속성을 이용합니다. 예를 들어 문장을 영어에서 프랑스어로 번역한 다음 프랑스어에서 영어로 다시 번역하면 원래 문장으로 돌아가야 한다는 의미입니다. [삼]. 수학적으로 번역기 G : X → Y와 다른 번역기 F : Y → X가 있는 경우 G와 F는 서로 역이고 두 매핑 모두 전단사여야 합니다. 매핑 G와 F를 동시에 훈련하고 F(G(x)) ≈ x 및 G(F(y)) ≈ y를 권장하는 주기 일관성 손실[64]을 추가하여 이 구조적 가정을 적용합니다. 이 손실을 도메인 X 및 Y에 대한 적대적 손실과 결합하면 짝을 이루지 않은 이미지 대 이미지 번역에 대한 완전한 목표를 얻을 수 있습니다.

We apply our method to a wide range of applications, including collection style transfer, object transfiguration, season transfer and photo enhancement. We also compare against previous approaches that rely either on hand-defined factorizations of style and content, or on shared embedding functions, and show that our method outperforms these baselines. We provide both PyTorch and Torch implementations. Check out more results at our website.

컬렉션 스타일 전송, 개체 변형, 시즌 전송 및 사진 보정을 포함한 광범위한 응용 프로그램에 우리의 방법을 적용합니다. 또한 스타일 및 콘텐츠의 손으로 정의한 인수분해 또는 공유 임베딩 기능에 의존하는 이전 접근 방식과 비교하여 우리의 방법이 이러한 기준선을 능가함을 보여줍니다. 우리는 PyTorch와 Torch 구현을 모두 제공합니다. 웹사이트에서 더 많은 결과를 확인하세요.

**2. Related work**

Generative Adversarial Networks (GANs) [16, 63] have achieved impressive results in image generation [6, 39], image editing [66], and representation learning [39, 43, 37]. Recent methods adopt the same idea for conditional image generation applications, such as text2image [41], image inpainting [38], and future prediction [36], as well as to other domains like videos [54] and 3D data [57]. The key to GANs’ success is the idea of an adversarial loss that forces the generated images to be, in principle, indistinguishable from real photos. This loss is particularly powerful for image generation tasks, as this is exactly the objective that much of computer graphics aims to optimize. We adopt an adversarial loss to learn the mapping such that the translated images cannot be distinguished from images in the target domain.

GAN(Generative Adversarial Networks)[16, 63]은 이미지 생성[6, 39], 이미지 편집[66] 및 표현 학습[39, 43, 37]에서 인상적인 결과를 달성했습니다. 최근의 방법은 text2image[41], image inpainting[38], 미래 예측[36]과 같은 조건부 이미지 생성 응용 프로그램과 비디오[54] 및 3D 데이터[57]와 같은 다른 영역에 대해 동일한 아이디어를 채택합니다. GAN 성공의 핵심은 생성된 이미지를 원칙적으로 실제 사진과 구별할 수 없도록 하는 적대적 손실에 대한 아이디어입니다. 이 손실은 이미지 생성 작업에 특히 강력합니다. 이것이 바로 많은 컴퓨터 그래픽이 최적화하는 목표이기 때문입니다. 번역된 이미지가 대상 도메인의 이미지와 구별될 수 없도록 매핑을 학습하기 위해 적대적 손실을 채택합니다.

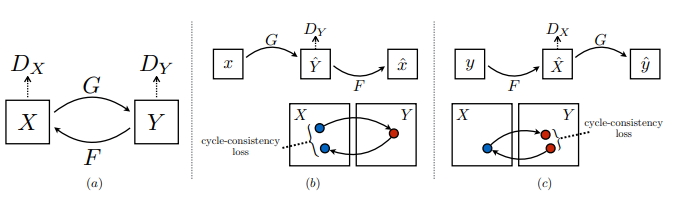


Figure 3: (a) Our model contains two mapping functions G : X → Y and F : Y → X, and associated adversarial discriminators and . encourages G to translate X into outputs indistinguishable from domain Y , and vice versa for and F. To further regularize the mappings, we introduce two cycle consistency losses that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: x → G(x) → F(G(x)) ≈ x, and (c) backward cycle-consistency loss: y → F(y) → G(F(y)) ≈ y

그림 3: (a) 우리 모델에는 두 개의 매핑 함수 G : X → Y 및 F : Y → X와 관련된 적대적 판별기 및 가 포함되어 있습니다. 는 G가 X를 도메인 Y와 구별할 수 없는 출력으로 변환하도록 장려하고, 와 F에 대해서도 그 반대의 경우도 마찬가지입니다. 매핑을 추가로 정규화하기 위해 우리는 한 도메인에서 다른 도메인으로 그리고 다시 역으로 변환할 경우 직관을 포착하는 두 개의 주기 일관성 손실을 도입합니다. (b) 순방향 주기 일관성 손실: x → G(x) → F(G(x)) ≈ x, (c) 역방향 주기 일관성 손실: y → F(y) → G(F(y)) ≈ y

Image-to-Image Translation The idea of image-toimage translation goes back at least to Hertzmann et al.’s Image Analogies [19], who employ a non-parametric texture model [10] on a single input-output training image pair. More recent approaches use a dataset of input-output examples to learn a parametric translation function using CNNs (e.g., [33]). Our approach builds on the “pix2pix” framework of Isola et al. [22], which uses a conditional generative adversarial network [16] to learn a mapping from input to output images. Similar ideas have been applied to various tasks such as generating photographs from sketches [44] or from attribute and semantic layouts [25]. However, unlike the above prior work, we learn the mapping without paired training examples.

Image-to-Image Translation 이미지-to-image 번역의 아이디어는 적어도 Hertzmann et al.의 Image Analogies[19]로 거슬러 올라가며, 이들은 단일 입력-출력 훈련 이미지 쌍에 비모수적 텍스처 모델[10]을 사용합니다. . 보다 최근의 접근 방식은 입력-출력 예제의 데이터 세트를 사용하여 CNN을 사용하여 매개변수 변환 기능을 학습합니다(예: [33]). 우리의 접근 방식은 Isola 등의 "pix2pix" 프레임워크를 기반으로 합니다. 입력 이미지에서 출력 이미지로의 매핑을 학습하기 위해 조건부 생성 적대 네트워크를 사용합니다. 스케치 또는 속성 및 의미론적 레이아웃에서 사진을 생성하는 것과 같은 다양한 작업에 유사한 아이디어가 적용되었습니다. 그러나 위의 선행 작업과 달리 paired training example 없이 mapping을 학습한다.

Unpaired Image-to-Image Translation Several other methods also tackle the unpaired setting, where the goal is to relate two data domains: X and Y . Rosales et al. [42] propose a Bayesian framework that includes a prior based on a patch-based Markov random field computed from a source image and a likelihood term obtained from multiple style images. More recently, CoGAN [32] and cross-modal scene networks [1] use a weight-sharing strategy to learn a common representation across domains. Concurrent to our method, Liu et al. [31] extends the above framework with a combination of variational autoencoders [27] and generative adversarial networks [16]. Another line of concurrent work [46, 49, 2] encourages the input and output to share specific “content” features even though they may differ in “style“. These methods also use adversarial networks, with additional terms to enforce the output to be close to the input in a predefined metric space, such as class label space [2], image pixel space [46], and image feature space [49].

짝을 이루지 않은 이미지에서 이미지로의 변환 여러 다른 방법도 짝을 이루지 않은 설정을 해결합니다. 여기서 목표는 X 와 Y 의 두 데이터 도메인을 연결하는 것입니다. Rosales et al. [42]는 소스 이미지에서 계산된 패치 기반 Markov 랜덤 필드와 다중 스타일 이미지에서 얻은 가능성 항을 기반으로 하는 사전을 포함하는 베이지안 프레임워크를 제안합니다. 보다 최근에는 CoGAN 및 교차 모드 장면 네트워크가 가중치 공유 전략을 사용하여 도메인 전체에서 공통 표현을 학습합니다. 우리의 방법과 동시에 Liu et al. [31]은 변형 자동 인코더와 생성적 적대 네트워크[16]의 조합으로 위의 프레임워크를 확장합니다. 동시 작업의 또 다른 라인[46, 49, 2]은 입력과 출력이 "스타일"이 다를지라도 특정 "콘텐츠" 기능을 공유하도록 권장합니다. 이러한 방법은 또한 클래스 레이블 공간[2], 이미지 픽셀 공간[46] 및 이미지 특징 공간[49]과 같은 미리 정의된 메트릭 공간에서 출력이 입력에 가깝게 되도록 하는 추가 용어와 함께 적대적 네트워크를 사용합니다.

Unlike the above approaches, our formulation does not rely on any task-specific, predefined similarity function between the input and output, nor do we assume that the input and output have to lie in the same low-dimensional embedding space. This makes our method a general-purpose solution for many vision and graphics tasks. We directly compare against several prior and contemporary approaches in Section 5.1.

위의 접근 방식과 달리 우리의 공식화는 입력과 출력 간의 사전 정의된 유사성 함수에 의존하지 않으며 입력과 출력이 동일한 저차원 임베딩 공간에 있어야 한다고 가정하지 않습니다. 이것은 우리의 방법을 많은 비전 및 그래픽 작업을 위한 범용 솔루션으로 만듭니다. 섹션 5.1에서 여러 이전 및 현대 접근 방식과 직접 비교합니다.

Cycle Consistency The idea of using transitivity as a way to regularize structured data has a long history. In visual tracking, enforcing simple forward-backward consistency has been a standard trick for decades [24, 48]. In the language domain, verifying and improving translations via “back translation and reconciliation” is a technique used by human translators [3] (including, humorously, by Mark Twain [51]), as well as by machines [17]. More recently, higher-order cycle consistency has been used in structure from motion [61], 3D shape matching [21], cosegmentation [55], dense semantic alignment [65, 64], and depth estimation [14]. Of these, Zhou et al. [64] and Godard et al. [14] are most similar to our work, as they use a cycle consistency loss as a way of using transitivity to supervise CNN training. In this work, we are introducing a similar loss to push G and F to be consistent with each other. Concurrent with our work, in these same proceedings, Yi et al. [59] independently use a similar objective for unpaired image-to-image translation, inspired by dual learning in machine translation [17].

주기 일관성 구조화된 데이터를 정규화하는 방법으로 전이성을 사용한다는 아이디어는 오랜 역사를 가지고 있습니다. 시각적 추적에서 간단한 앞뒤 일관성을 적용하는 것은 수십 년 동안 표준 트릭이었습니다. 언어 영역에서 "역 번역 및 조정"을 통해 번역을 확인하고 개선하는 것은 기계뿐만 아니라 인간 번역가(Mark Twai도 유머러스하게 포함)에서 사용하는 기술입니다. 보다 최근에는 고차 주기 일관성이 움직임으로부터의 구조[61], 3D 형상 일치, 코세그멘테이션, 조밀한 의미적 정렬 및 깊이 추정[14]에서 사용되었습니다. 이 중 Zhou et al. [64] 및 Godard et al. [14] CNN 훈련을 감독하기 위해 전이성을 사용하는 방법으로 주기 일관성 손실을 사용하기 때문에 우리 작업과 가장 유사합니다. 이 작업에서는 G와 F를 서로 일치시키기 위해 유사한 손실을 도입하고 있습니다. 우리 작업과 동시에 이러한 동일한 절차에서 Yi et al. [59] 기계 번역의 이중 학습에서 영감을 받아 짝을 이루지 않은 이미지 대 이미지 번역에 대해 유사한 목표를 독립적으로 사용합니다.

Neural Style Transfer [13, 23, 52, 12] is another way to perform image-to-image translation, which synthesizes a novel image by combining the content of one image with the style of another image (typically a painting) based on matching the Gram matrix statistics of pre-trained deep features. Our primary focus, on the other hand, is learning the mapping between two image collections, rather than between two specific images, by trying to capture correspondences between higher-level appearance structures. Therefore, our method can be applied to other tasks, such as painting→ photo, object transfiguration, etc. where single sample transfer methods do not perform well. We compare these two methods in Section 5.2.

Neural Style Transf는 매칭을 기반으로 한 이미지의 내용을 다른 이미지(일반적으로 그림)의 스타일과 결합하여 새로운 이미지를 합성하는 이미지 대 이미지 번역을 수행하는 또 다른 방법입니다. 사전 훈련된 심층 기능의 그램 행렬 통계입니다. 반면에 우리의 주요 초점은 더 높은 수준의 외관 구조 간의 대응을 포착하려고 시도함으로써 두 개의 특정 이미지가 아니라 두 개의 이미지 컬렉션 간의 매핑을 학습하는 것입니다. 따라서 우리의 방법은 단일 샘플 전송 방법이 잘 수행되지 않는 페인팅 → 사진, 개체 변형 등과 같은 다른 작업에 적용할 수 있습니다. 섹션 5.2에서 이 두 가지 방법을 비교합니다.

**3. Formulation**

Our goal is to learn mapping functions between two domains X and Y given training samples where and where . We denote the data distribution as x ~ and y ~ . As illustrated in Figure 3 (a), our model includes two mappings G : X Y and F : Y X. In addition, we introduce two adversarial discriminators and , where DX aims to distinguish between images {x} and translated images {F(y)}; in the same way, aims to discriminate between {y} and {G(x)}. Our objective contains two types of terms: adversarial losses [16] for matching the distribution of generated images to the data distribution in the target domain; and cycle consistency losses to prevent the learned mappings G and F from contradicting each other.

우리의 목표는 주어진 훈련 샘플 에서 , 에서 에서 두 도메인 X와 Y 사이의 매핑 함수를 배우는 것입니다. 데이터 분포를 x ~ 및 y ~ 로 나타냅니다. 그림 3 (a)에서 설명한 바와 같이, 우리의 모델에는 두 개의 매핑 G: X → Y 및 F: Y → X가 포함됩니다. 또한, 우리는 두 개의 적대적 식별자 및 를 소개합니다. 여기서 DX는 이미지 {x}과 변환된 이미지 {F(y)를 구별하는 것을 목표로 하며, 는 동일한 방식으로 {x}과 식별을 목표로 합니다.생성된 영상의 분포를 대상 도메인의 데이터 분포와 일치시키기 위한 Versarial 손실[16] 및 학습된 매핑 G와 F가 서로 모순되지 않도록 일관성 손실을 순환합니다.

**3.1. Adversarial Loss**

We apply adversarial losses [16] to both mapping functions. For the mapping function G : X Y and its discriminator , we express the objective as:

우리는 두 매핑 기능에 적대적 손실[16]을 적용합니다. 매핑 함수 G : X → Y 및 판별자 D\_y 의 경우 목적을 다음과 같이 표현합니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

where G tries to generate images G(x) that look similar to images from domain Y , while aims to distinguish between translated samples G(x) and real samples y. G aims to minimize this objective against an adversary D that tries to maximize it, i.e., .We introduce a similar adversarial loss for the mapping function F : Y X and its discriminator DX as well: i.e., .

여기서 G는 도메인 Y의 이미지와 유사한 이미지 G(x)를 생성하려고 하는 반면, 는 변환된 샘플 G(x)와 실제 샘플 y를 구별하려고 합니다. G는 최대화를 시도하는 상대 D, 즉 에 대해 이러한 목표를 최소화하는 것을 목표로 합니다.매핑 함수 F : Y → X 및 해당 식별자 DX에도 유사한 적대적 손실을 도입합니다. 즉, 입니다.

**3.2. Cycle Consistency Loss**

Adversarial training can, in theory, learn mappings G and F that produce outputs identically distributed as target domains Y and X respectively (strictly speaking, this requires G and F to be stochastic functions) [15]. However, with large enough capacity, a network can map the same set of input images to any random permutation of images in the target domain, where any of the learned mappings can induce an output distribution that matches the target distribution. Thus, adversarial losses alone cannot guarantee that the learned function can map an individual input to a desired output . To further reduce the space of possible mapping functions, we argue that the learned mapping functions should be cycle-consistent: as shown in Figure 3 (b), for each image x from domain X, the image translation cycle should be able to bring x back to the original image, i.e.,. We call this forward cycle consistency. Similarly, as illustrated in Figure 3 (c), for each image y from domain Y , G and F should also satisfy backward cycle consistency:. We incentivize this behavior using a cycle consistency loss:

적대적 훈련은 이론적으로 각각 목표 영역 Y와 X와 동일하게 분포된 출력을 생성하는 매핑 G와 F를 학습할 수 있습니다 (엄밀히 말하면, 이것은 G와 F가 확률 함수여야 합니다) [15]. 그러나 용량이 충분히 크면 네트워크는 대상 도메인에 있는 이미지의 임의 배열에 동일한 입력 이미지 집합을 매핑할 수 있습니다. 여기서 학습된 매핑은 대상 분포와 일치하는 출력 분포를 유도할 수 있습니다. 그러므로, 적대적 손실만으로는 학습된 함수가 개별 입력 를 원하는 출력 에 매핑할 수 있다는 것을 보장할 수 없습니다. 가능한 매핑 기능의 공간을 추가로 줄이기 위해 학습된 매핑 함수는 주기 정합성이 있어야 한다고 주장합니다. 그림 3 (b)와 같이 도메인 X의 각 이미지 x에 대해 이미지 변환 주기는 x를 원래 이미지로 되돌릴 수 있어야 합니다. 즉, 입니다. 이를 전진 주기 일관성이라고 합니다. 그림 3 (c)에서 설명한 것과 같이, 도메인 Y, G 및 F의 각 이미지 y에 대해서도 역방향 주기 일관성을 충족해야 합니다. . clossy 주기 일관성 손실을 사용하여 이 동작을 장려합니다.

텍스트, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 4: The input images x, output images G(x) and the reconstructed images F(G(x)) from various experiments. From top to bottom: photo↔Cezanne, horses↔zebras, winter→summer Yosemite, aerial photos↔Google maps.

그림 4: 입력 이미지 x, 출력 이미지 G(x) 및 다양한 실험에서 재구성된 이미지 F(G(x)). 위에서 아래로 사진↔세잔느, 말↔얼룩말, 겨울→여름 요세미티, 항공 사진↔구글 지도.

텍스트이(가) 표시된 사진

자동 생성된 설명

In preliminary experiments, we also tried replacing the L1 norm in this loss with an adversarial loss between F(G(x)) and x, and between G(F(y)) and y, but did not observe improved performance. The behavior induced by the cycle consistency loss can be observed in Figure 4: the reconstructed images F(G(x)) end up matching closely to the input images x.

예비 실험에서 우리는 이 손실의 L1 노름을 F(G(x))와 x 사이, G(F(y))와 y 사이의 적대적 손실로 대체하려고 시도했지만 개선된 성능을 관찰하지 못했습니다. 주기 일관성 손실에 의해 유도된 동작은 그림 4에서 관찰할 수 있습니다. 재구성된 이미지 F(G(x))는 입력 이미지 x와 밀접하게 일치하게 됩니다.

**3.3. Full Objective**

Our full objective is:

우리의 완전한 목표는 다음과 같습니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

where controls the relative importance of the two objectives. We aim to solve:

여기서 λ는 두 목표의 상대적 중요성을 제어합니다. 우리는 다음을 해결하는 것을 목표로 합니다.



Notice that our model can be viewed as training two “autoencoders” [20]: we learn one autoencoder F G : X X jointly with another G F : Y Y . However, these autoencoders each have special internal structures: they map an image to itself via an intermediate representation that is a translation of the image into another domain. Such a setup can also be seen as a special case of “adversarial autoencoders” [34], which use an adversarial loss to train the bottleneck layer of an autoencoder to match an arbitrary tar- get distribution. In our case, the target distribution for the X X autoencoder is that of the domain Y .

우리 모델은 두 개의 "자동 인코더"[20]를 훈련하는 것으로 볼 수 있습니다. 하나의 자동 인코더 F ° G : X →X를 다른 G ° F : Y→ Y와 함께 학습합니다. 그러나 이러한 자동 인코더에는 각각 특별한 내부 구조가 있습니다. 즉, 이미지를 다른 도메인으로 변환하는 중간 표현을 통해 이미지를 자체적으로 매핑합니다. 이러한 설정은 적대적 손실을 사용하여 임의의 대상 분포와 일치하도록 자동 인코더의 병목 계층을 훈련시키는 "적대적 자동 인코더"[34]의 특별한 경우로도 볼 수 있습니다. 우리의 경우 X→ X 자동 인코더의 대상 분포는 도메인 Y의 분포입니다.

In Section 5.1.4, we compare our method against ablations of the full objective, including the adversarial loss LGAN alone and the cycle consistency loss alone, and empirically show that both objectives play critical roles in arriving at high-quality results. We also evaluate our method with only cycle loss in one direction and show that a single cycle is not sufficient to regularize the training for this under-constrained problem.

섹션 5.1.4에서 우리의 방법을 적대적 손실 LGAN 단독과 주기 일관성 손실 단독을 포함하여 전체 목표의 제거와 비교하고 두 목표가 고품질 결과에 도달하는 데 중요한 역할을 한다는 것을 경험적으로 보여줍니다. 우리는 또한 한 방향으로만 사이클 손실을 사용하여 방법을 평가하고 단일 사이클이 이 under-constrained 문제에 대한 훈련을 정규화하기에 충분하지 않다는 것을 보여줍니다.

**4. Implementation**

Network Architecture We adopt the architecture for our generative networks from Johnson et al. [23] who have shown impressive results for neural style transfer and superresolution. This network contains three convolutions, several residual blocks [18], two fractionally-strided convolutions with stride 1 2 , and one convolution that maps features to RGB. We use 6 blocks for 128 × 128 images and 9 blocks for 256×256 and higher-resolution training images. Similar to Johnson et al. [23], we use instance normalization [53]. For the discriminator networks we use 70 × 70 PatchGANs [22, 30, 29], which aim to classify whether 70 × 70 overlapping image patches are real or fake. Such a patch-level discriminator architecture has fewer parameters than a full-image discriminator and can work on arbitrarilysized images in a fully convolutional fashion [22].

네트워크 아키텍처 우리는 Johnson et al.의 생성 네트워크에 대한 아키텍처를 채택합니다. [23] 신경 스타일 전달 및 초해상도에 대해 인상적인 결과를 보여주었습니다. 이 네트워크에는 3개의 컨볼루션, 여러 잔차 블록[18], 보폭이 1 2인 2개의 fractionally-strided convolution 및 기능을 RGB에 매핑하는 1개의 컨볼루션이 포함됩니다. 128×128 이미지에는 6개의 블록을, 256×256 이상의 고해상도 훈련 이미지에는 9개의 블록을 사용합니다. Johnson et al.과 유사합니다. [23], 우리는 인스턴스 정규화 [53]를 사용합니다. 판별자 네트워크의 경우 70 × 70 PatchGAN[22, 30, 29]을 사용하여 70 × 70 겹치는 이미지 패치가 진짜인지 가짜인지 분류하는 것을 목표로 합니다. 이러한 패치 수준 판별기 아키텍처는 전체 이미지 판별기보다 매개변수가 적고 완전한 컨볼루션 방식으로 임의 크기의 이미지에 대해 작업할 수 있습니다[22].

Training details We apply two techniques from recent works to stabilize our model training procedure. First, for (Equation 1), we replace the negative log likelihood objective by a least-squares loss [35]. This loss is more stable during training and generates higher quality results. In particular, for a GAN loss (G, D, X, Y ), we train the G to minimize and train the D to minimize.

훈련 세부 사항 우리는 모델 훈련 절차를 안정화하기 위해 최근 작업의 두 가지 기술을 적용합니다. 먼저 (방정식 1)의 경우 음수 로그 가능성 목표를 최소 자승 손실로 대체합니다[35]. 이 손실은 훈련 중에 더 안정적이며 더 높은 품질의 결과를 생성합니다. 특히, GAN 손실 (G, D, X, Y )에 대해 를 최소화하도록 G를 훈련하고 D를 최소화하기 위해.

Second, to reduce model oscillation [15], we follow Shrivastava et al.’s strategy [46] and update the discriminators using a history of generated images rather than the ones produced by the latest generators. We keep an image buffer that stores the 50 previously created images. For all the experiments, we set in Equation 3. We use the Adam solver [26] with a batch size of 1. All networks were trained from scratch with a learning rate of 0.0002. We keep the same learning rate for the first 100 epochs and linearly decay the rate to zero over the next 100 epochs. Please see the appendix (Section 7) for more details about the datasets, architectures, and training procedures.

둘째, 모델 진동을 줄이기 위해 [15] Shrivastava 등의 전략 [46]을 따르고 최신 생성기에서 생성된 이미지가 아닌 생성된 이미지의 이력을 사용하여 판별자를 업데이트합니다. 우리는 이전에 생성된 50개의 이미지를 저장하는 이미지 버퍼를 유지합니다. 모든 실험에 대해 식 3에서 λ=10을 설정했습니다. 배치 크기가 1인 Adam 솔버[26]를 사용합니다. 모든 네트워크는 학습률 0.0002로 처음부터 훈련되었습니다. 처음 100개의 epoch에 대해 동일한 학습률을 유지하고 다음 100개의 epoch에 대해 비율을 선형적으로 0으로 감소시킵니다. 데이터 세트, 아키텍처 및 교육 절차에 대한 자세한 내용은 부록(섹션 7)을 참조하십시오.

**5. Results**

We first compare our approach against recent methods for unpaired image-to-image translation on paired datasets where ground truth input-output pairs are available for evaluation. We then study the importance of both the adversarial loss and the cycle consistency loss and compare our full method against several variants. Finally, we demonstrate the generality of our algorithm on a wide range of applications where paired data does not exist. For brevity, we refer to our method as CycleGAN. The PyTorch and Torch code, models, and full results can be found at our website.

먼저 우리의 접근 방식을 실측 입력-출력 쌍을 평가에 사용할 수 있는 쌍을 이루는 데이터 세트에서 쌍을 이루지 않은 이미지-이미지 변환을 위한 최근 방법과 비교합니다. 그런 다음 적대적 손실과 주기 일관성 손실의 중요성을 연구하고 전체 방법을 여러 변형과 비교합니다. 마지막으로 쌍 데이터가 존재하지 않는 광범위한 응용 프로그램에서 알고리즘의 일반성을 보여줍니다. 간결함을 위해 우리의 방법을 CycleGAN이라고 합니다. PyTorch 및 Torch 코드, 모델 및 전체 결과는 당사 웹사이트에서 찾을 수 있습니다.

**5.1. Evaluation**

Using the same evaluation datasets and metrics as “pix2pix” [22], we compare our method against several baselines both qualitatively and quantitatively. The tasks include semantic labels↔photo on the Cityscapes dataset [4], and map↔aerial photo on data scraped from Google Maps. We also perform ablation study on the full loss function.

"pix2pix"[22]와 동일한 평가 데이터 세트 및 메트릭을 사용하여 우리의 방법을 질적 및 양적으로 여러 기준선과 비교합니다. 작업에는 Cityscapes 데이터세트[4]의 의미적 레이블↔사진과 Google 지도에서 스크랩한 데이터의 지도↔항공 사진이 포함됩니다. 또한 전체 손실 함수에 대한 절제 연구도 수행합니다.

**5.1.1 Evaluation Metrics**

AMT perceptual studies On the map↔aerial photo task, we run “real vs fake” perceptual studies on Amazon Mechanical Turk (AMT) to assess the realism of our outputs. We follow the same perceptual study protocol from Isola et al. [22], except we only gather data from 25 participants per algorithm we tested. Participants were shown a sequence of pairs of images, one a real photo or map and one fake (generated by our algorithm or a baseline), and asked to click on the image they thought was real. The first 10 trials of each session were practice and feedback was given as to whether the participant’s response was correct or incorrect. The remaining 40 trials were used to assess the rate at which each algorithm fooled participants. Each session only tested a single algorithm, and participants were only allowed to complete a single session. The numbers we report here are not directly comparable to those in [22] as our ground truth images were processed slightly differently 2 and the participant pool we tested may be differently distributed from those tested in [22] (due to running the experiment at a different date and time). Therefore, our numbers should only be used to compare our current method against the baselines (which were run under identical conditions), rather than against [22].

AMT 지각 연구 지도↔항공 사진 작업에서 Amazon Mechanical Turk(AMT)에 대한 "실제 대 가짜" 지각 연구를 실행하여 출력의 사실감을 평가합니다. 우리는 Isola 등의 동일한 지각 연구 프로토콜을 따릅니다. [22], 우리가 테스트한 알고리즘당 25명의 참가자로부터 데이터만 수집한다는 점을 제외하고. 참가자들에게 실제 사진 또는 지도와 가짜(저희 알고리즘 또는 기준선에 의해 생성됨)로 구성된 일련의 이미지 쌍을 보여주고 실제라고 생각하는 이미지를 클릭하도록 요청했습니다. 각 세션의 처음 10번의 시도는 연습이었고 참가자의 응답이 올바른지 잘못된지에 대한 피드백이 제공되었습니다. 나머지 40개 시도는 각 알고리즘이 참가자를 속인 비율을 평가하는 데 사용되었습니다. 각 세션은 단일 알고리즘만 테스트했으며 참가자는 단일 세션만 완료할 수 있었습니다. 우리가 여기에 보고한 수치는 우리의 실측 이미지가 약간 다르게 처리되었고 2 우리가 테스트한 참가자 풀이 [22]에서 테스트된 것과 다르게 분포될 수 있기 때문에 [22]의 수치와 직접적으로 비교할 수 없습니다(실험을 [22]에서 실행하기 때문에 다른 날짜와 시간). 따라서 우리의 수치는 [22]가 아니라 기준선(동일한 조건에서 실행된)과 현재 방법을 비교하는 데에만 사용해야 합니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 5: Different methods for mapping labels↔photos trained on Cityscapes images. From left to right: input, BiGAN/ALI [7, 9], CoGAN [32], feature loss + GAN, SimGAN [46], CycleGAN (ours), pix2pix [22] trained on paired data, and ground truth.

그림 5: Cityscapes 이미지에서 훈련된 레이블↔사진을 매핑하는 다양한 방법. 왼쪽에서 오른쪽으로: 입력, BiGAN/ALI [7, 9], CoGAN [32], 기능 손실 + GAN, SimGAN [46], CycleGAN(당사), 짝지어진 데이터에 대해 훈련된 pix2pix [22] 및 정답.

텍스트, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 6: Different methods for mapping aerial photos↔maps on Google Maps. From left to right: input, BiGAN/ALI [7, 9], CoGAN [32], feature loss + GAN, SimGAN [46], CycleGAN (ours), pix2pix [22] trained on paired data, and ground truth.

그림 6: Google 지도에서 항공 사진↔지도를 매핑하는 다양한 방법. 왼쪽에서 오른쪽으로: 입력, BiGAN/ALI [7, 9], CoGAN [32], 기능 손실 + GAN, SimGAN [46], CycleGAN(당사), 짝지어진 데이터에 대해 훈련된 pix2pix [22] 및 정답.

FCN score Although perceptual studies may be the gold standard for assessing graphical realism, we also seek an automatic quantitative measure that does not require human experiments. For this, we adopt the “FCN score” from [22], and use it to evaluate the Cityscapes labels→photo task. The FCN metric evaluates how interpretable the generated photos are according to an off-the-shelf semantic segmentation algorithm (the fully-convolutional network, FCN, from [33]). The FCN predicts a label map for a generated photo. This label map can then be compared against the input ground truth labels using standard semantic segmentation metrics described below. The intuition is that if we generate a photo from a label map of “car on the road”, then we have succeeded if the FCN applied to the generated photo detects “car on the road”.

FCN 점수 지각적 연구가 그래픽 리얼리즘을 평가하기 위한 금본위제일 수 있지만, 우리는 또한 인간의 실험을 필요로 하지 않는 자동 정량적 측정을 추구합니다. 이를 위해 [22]의 "FCN 점수"를 채택하여 Cityscapes 레이블 → 사진 작업을 평가하는 데 사용합니다. FCN 메트릭은 기성 시맨틱 분할 알고리즘([33]의 완전 컨볼루션 네트워크, FCN)에 따라 생성된 사진이 얼마나 해석 가능한지 평가합니다. FCN은 생성된 사진에 대한 레이블 맵을 예측합니다. 그런 다음 이 레이블 맵은 아래에 설명된 표준 의미론적 세분화 메트릭을 사용하여 입력 정답 레이블과 비교할 수 있습니다. 직관은 "도로 위의 자동차"라는 레이블 맵에서 사진을 생성하면 생성된 사진에 적용된 FCN이 "도로 위의 자동차"를 감지하면 성공한 것입니다.

Semantic segmentation metrics To evaluate the performance of photo→labels, we use the standard metrics from the Cityscapes benchmark [4], including per-pixel accuracy, per-class accuracy, and mean class Intersection-Over-Union (Class IOU) [4].

의미론적 세분화 메트릭 사진→라벨의 성능을 평가하기 위해 픽셀당 정확도, 클래스당 정확도, 평균 클래스 Intersection-Over-Union(Class IOU)을 포함하여 Cityscapes 벤치마크[4]의 표준 메트릭을 사용합니다. ].

**5.1.2 Baselines**

CoGAN [32] This method learns one GAN generator for domain X and one for domain Y , with tied weights on the first few layers for shared latent representations. Translation from X to Y can be achieved by finding a latent representation that generates image X and then rendering this latent representation into style Y .

CoGAN [32] 이 방법은 공유 잠재 표현을 위해 처음 몇 개의 레이어에 가중치를 부여하여 도메인 X와 도메인 Y에 대해 하나의 GAN 생성기를 학습합니다. X에서 Y로의 변환은 이미지 X를 생성하는 잠재 표현을 찾은 다음 이 잠재 표현을 스타일 Y로 렌더링하여 달성할 수 있습니다.

SimGAN [46] Like our method, Shrivastava et al.[46] uses an adversarial loss to train a translation from X to Y . The regularization term is used to penalize making large changes at pixel level.

SimGAN [46] 우리의 방법과 마찬가지로 Shrivastava et al.[46] X 에서 Y 로의 번역을 훈련하기 위해 적대적 손실을 사용합니다. 정규화 항 |x-G(x)| 픽셀 수준에서 큰 변경을 벌일 때 벌점을 받는 데 사용됩니다.

텍스트, 테이블이(가) 표시된 사진

자동 생성된 설명

Table 1: AMT “real vs fake” test on maps↔aerial photos at 256 × 256 resolution.

표 1: 256 × 256 해상도의 지도↔항공 사진에 대한 AMT "진짜 대 가짜" 테스트.

테이블이(가) 표시된 사진

자동 생성된 설명

Table 2: FCN-scores for different methods, evaluated on Cityscapes labels→photo.

표 2: Cityscapes 레이블→사진에서 평가된 다양한 방법에 대한 FCN 점수.

테이블이(가) 표시된 사진

자동 생성된 설명

Table 3: Classification performance of photo→labels for different methods on cityscapes.

표 3: 도시 경관에 대한 다양한 방법에 대한 사진→라벨의 분류 성능.

Feature loss + GAN We also test a variant of SimGAN [46] where the L1 loss is computed over deep image features using a pretrained network (VGG-16 relu4 2 [47]), rather than over RGB pixel values. Computing distances in deep feature space, like this, is also sometimes referred to as using a “perceptual loss” [8, 23].

특징 손실 + GAN 또한 RGB 픽셀 값이 아닌 사전 훈련된 네트워크(VGG-16 relu4 2 [47])를 사용하여 깊은 이미지 특징에 대해 L1 손실이 계산되는 SimGAN의 변형을 테스트합니다. 이와 같이 깊은 피쳐 공간에서 거리를 계산하는 것은 때때로 "지각 손실"을 사용하는 것으로도 언급됩니다[8, 23].

BiGAN/ALI [9, 7] Unconditional GANs [16] learn a generator G : Z → X, that maps a random noise z to an image x. The BiGAN [9] and ALI [7] propose to also learn the inverse mapping function F : X → Z. Though they were originally designed for mapping a latent vector z to an image x, we implemented the same objective for mapping a source image x to a target image y.

BiGAN/ALI [9, 7] 무조건 GAN[16]은 랜덤 노이즈 z를 이미지 x에 매핑하는 생성기 G : Z → X를 학습합니다. BiGAN[9]과 ALI[7]는 역 매핑 함수 F : X → Z도 학습할 것을 제안합니다. 원래 잠재 벡터 z를 이미지 x에 매핑하도록 설계되었지만 소스 이미지 매핑에 대해 동일한 목표를 구현했습니다. x를 대상 이미지 y로.

pix2pix [22] We also compare against pix2pix [22], which is trained on paired data, to see how close we can get to this “upper bound” without using any paired data.

pix2pix [22] 또한 쌍 데이터에 대해 훈련된 pix2pix[22]와 비교하여 쌍 데이터를 사용하지 않고 이 "상한"에 얼마나 근접할 수 있는지 확인합니다.

For a fair comparison, we implement all the baselines using the same architecture and details as our method, except for CoGAN [32]. CoGAN builds on generators that produce images from a shared latent representation, which is incompatible with our image-to-image network. We use the public implementation of CoGAN instead

공정한 비교를 위해 CoGAN [32]을 제외하고 우리 방법과 동일한 아키텍처와 세부 사항을 사용하여 모든 기준선을 구현합니다. CoGAN은 우리의 이미지 대 이미지 네트워크와 호환되지 않는 공유 잠재 표현에서 이미지를 생성하는 생성기를 기반으로 합니다. 대신 CoGAN의 공개 구현을 사용합니다.

**5.1.3 Comparison against baselines**

As can be seen in Figure 5 and Figure 6, we were unable to achieve compelling results with any of the baselines. Our method, on the other hand, can produce translations that are often of similar quality to the fully supervised pix2pix.

그림 5와 그림 6에서 볼 수 있듯이 어떤 기준으로도 설득력 있는 결과를 얻을 수 없었습니다. 반면에 우리의 방법은 완전히 감독된 pix2pix와 비슷한 품질의 번역을 생성할 수 있습니다.

테이블이(가) 표시된 사진

자동 생성된 설명

Table 4: Ablation study: FCN-scores for different variants of our method, evaluated on Cityscapes labels→photo.

표 4: 절제 연구: 우리 방법의 다양한 변형에 대한 FCN 점수, Cityscapes 레이블 → 사진에서 평가됨.

테이블이(가) 표시된 사진

자동 생성된 설명

Table 5: Ablation study: classification performance of photo→labels for different losses, evaluated on Cityscapes.

표 5: 절제 연구: 사진의 분류 성능 → 다양한 손실에 대한 레이블, Cityscapes에서 평가됨.

Table 1 reports performance regarding the AMT perceptual realism task. Here, we see that our method can fool participants on around a quarter of trials, in both the maps→aerial photos direction and the aerial photos→maps direction at 256 × 256 resolution3 . All the baselines almost never fooled participants.

표 1은 AMT 지각 현실주의 작업에 대한 성능을 보고합니다. 여기에서 우리의 방법이 256 × 256 해상도에서 지도→항공 사진 방향 및 항공 사진→지도 방향 모두에서 약 4분의 1의 실험에서 참가자를 속일 수 있음을 알 수 있습니다. 모든 기준선은 참가자를 거의 속이지 않았습니다.

Table 2 assesses the performance of the labels→photo task on the Cityscapes and Table 3 evaluates the opposite mapping (photos→labels). In both cases, our method again outperforms the baselines.

표 2는 도시경관에 대한 레이블→사진 작업의 성능을 평가하고 표 3은 반대 매핑(사진→레이블)을 평가합니다. 두 경우 모두 우리의 방법은 다시 기준선을 능가합니다.

**5.1.4 Analysis of the loss function**

In Table 4 and Table 5, we compare against ablations of our full loss. Removing the GAN loss substantially degrades results, as does removing the cycle-consistency loss. We therefore conclude that both terms are critical to our results. We also evaluate our method with the cycle loss in only one direction: GAN + forward cycle loss , or GAN + backward cycle loss (Equation 2) and find that it often incurs training instability and causes mode collapse, especially for the direction of the mapping that was removed. Figure 7 shows several qualitative examples.

표 4와 표 5에서 우리는 전체 손실의 절제와 비교합니다. GAN 손실을 제거하면 주기 일관성 손실을 제거하는 것과 마찬가지로 결과가 크게 저하됩니다. 따라서 우리는 두 용어가 우리의 결과에 중요하다고 결론지었습니다. 또한 한 방향으로만 사이클 손실을 사용하여 방법을 평가합니다. GAN + 순방향 사이클 손실 또는 GAN + 역방향 사이클 손실 (수식 2) 특히 제거된 매핑 방향에 대해 훈련 불안정을 유발하고 모드 붕괴를 유발한다는 것을 발견했습니다. 몇 가지 질적 예를 보여줍니다.

**5.1.5 Image reconstruction quality**

In Figure 4, we show a few random samples of the reconstructed images F(G(x)). We observed that the reconstructed images were often close to the original inputs x, at both training and testing time, even in cases where one domain represents significantly more diverse information, such as map↔aerial photos.

그림 4에서 재구성된 이미지 F(G(x))의 몇 가지 무작위 샘플을 보여줍니다. 우리는 한 도메인이 지도↔항공 사진과 같이 훨씬 더 다양한 정보를 나타내는 경우에도 훈련 및 테스트 시간 모두에서 재구성된 이미지가 종종 원래 입력 x에 가깝다는 것을 관찰했습니다.

텍스트, 다른, 스크린샷, 다채로운이(가) 표시된 사진

자동 생성된 설명

Figure 7: Different variants of our method for mapping labels↔photos trained on cityscapes. From left to right: input, cycleconsistency loss alone, adversarial loss alone, GAN + forward cycle-consistency loss (F(G(x)) ≈ x), GAN + backward cycle-consistency loss (G(F(y)) ≈ y), CycleGAN (our full method), and ground truth. Both Cycle alone and GAN + backward fail to produce images similar to the target domain. GAN alone and GAN + forward suffer from mode collapse, producing identical label maps regardless of the input photo.

그림 7: 도시 경관에서 훈련된 라벨↔사진 매핑 방법의 다양한 변형. 왼쪽에서 오른쪽으로: 입력, 주기 일관성 손실만, 적대적 손실만, GAN + 순방향 주기-일관성 손실(F(G(x)) ≈ x), GAN + 역방향 주기-일관성 손실(G(F(y)) ≈ y), CycleGAN(우리의 전체 방법) 및 ground truth. Cycle 단독 및 GAN + 역방향 모두 대상 도메인과 유사한 이미지를 생성하지 못합니다. GAN 단독 및 GAN + forward는 모드 붕괴로 인해 입력 사진에 관계없이 동일한 레이블 맵을 생성합니다.

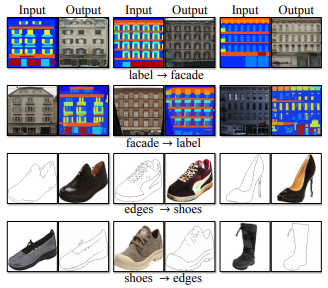


Figure 8: Example results of CycleGAN on paired datasets used in “pix2pix” [22] such as architectural labels↔photos and edges↔shoes.

그림 8: 건축 라벨↔사진 및 가장자리↔신발과 같이 "pix2pix"[22]에서 사용되는 쌍으로 된 데이터 세트에 대한 CycleGAN의 예시 결과.

**5.1.6 Additional results on paired datasets**

Figure 8 shows some example results on other paired datasets used in “pix2pix” [22], such as architectural labels↔photos from the CMP Facade Database [40], and edges↔shoes from the UT Zappos50K dataset [60]. The image quality of our results is close to those produced by the fully supervised pix2pix while our method learns the mapping without paired supervision

그림 8은 "pix2pix"[22]에 사용된 다른 쌍을 이루는 데이터 세트에 대한 몇 가지 예시 결과를 보여줍니다. 예를 들어 CMP Facade Database[40]의 건축 라벨↔사진, UT Zappos50K 데이터 세트[60]의 edge↔신발. 결과의 이미지 품질은 완전히 감독된 pix2pix에 의해 생성된 것과 비슷하지만 우리의 방법은 쌍을 이루는 감독 없이 매핑을 학습합니다.

**5.2. Applications**

We demonstrate our method on several applications where paired training data does not exist. Please refer to the appendix (Section 7) for more details about the datasets. We observe that translations on training data are often more appealing than those on test data, and full results of all applications on both training and test data can be viewed on our project website.

우리는 쌍을 이루는 훈련 데이터가 존재하지 않는 여러 애플리케이션에서 우리의 방법을 시연합니다. 데이터 세트에 대한 자세한 내용은 부록(섹션 7)을 참조하십시오. 우리는 훈련 데이터에 대한 번역이 테스트 데이터에 대한 번역보다 종종 더 매력적이며 훈련 및 테스트 데이터에 대한 모든 응용 프로그램의 전체 결과를 프로젝트 웹 사이트에서 볼 수 있음을 관찰했습니다.

Collection style transfer (Figure 10 and Figure 11) We train the model on landscape photographs downloaded from Flickr and WikiArt. Unlike recent work on “neural style transfer” [13], our method learns to mimic the style of an entire collection of artworks, rather than transferring the style of a single selected piece of art. Therefore, we can learn to generate photos in the style of, e.g., Van Gogh, rather than just in the style of Starry Night. The size of the dataset for each artist/style was 526, 1073, 400, and 563 for Cezanne, Monet, Van Gogh, and Ukiyo-e.

컬렉션 스타일 전송(그림 10 및 그림 11) Flickr 및 WikiArt에서 다운로드한 풍경 사진으로 모델을 훈련합니다. "신경 스타일 전이"[13]에 대한 최근 작업과 달리 우리의 방법은 선택된 단일 예술 작품의 스타일을 전송하는 것이 아니라 전체 예술 컬렉션의 스타일을 모방하는 방법을 배웁니다. 따라서 우리는 별이 빛나는 밤 스타일이 아니라 예를 들어 반 고흐 스타일로 사진을 생성하는 방법을 배울 수 있습니다. 각 아티스트/스타일에 대한 데이터 세트의 크기는 Cezanne, Monet, Van Gogh 및 Ukiyo-e의 경우 526, 1073, 400 및 563이었습니다.

Object transfiguration (Figure 13) The model is trained to translate one object class from ImageNet [5] to another (each class contains around 1000 training images). Turmukhambetov et al. [50] propose a subspace model to translate one object into another object of the same category, while our method focuses on object transfiguration between two visually similar categories.

객체 변형(그림 13) 이 모델은 ImageNet[5]의 한 객체 클래스를 다른 객체 클래스로 변환하도록 훈련되었습니다(각 클래스에는 약 1000개의 훈련 이미지가 포함됨). Turmukhambetov et al. [50]은 한 객체를 동일한 범주의 다른 객체로 변환하는 부분 공간 모델을 제안하는 반면, 우리의 방법은 시각적으로 유사한 두 범주 간의 객체 변형에 중점을 둡니다.

Season transfer (Figure 13) The model is trained on 854 winter photos and 1273 summer photos of Yosemite downloaded from Flickr.

시즌 전환(그림 13) 모델은 Flickr에서 다운로드한 요세미티의 겨울 사진 854장과 여름 사진 1273장으로 훈련되었습니다.

Photo generation from paintings (Figure 12) For paintingphoto, we find that it is helpful to introduce an additional loss to encourage the mapping to preserve color composition between the input and output. In particular, we adopt the technique of Taigman et al. [49] and regularize the generator to be near an identity mapping when real samples of the target domain are provided as the input to the generator: i.e.,

그림에서 사진 생성(그림 12) painting→사진의 경우 입력과 출력 사이의 색 구성을 보존하기 위해 매핑을 장려하기 위해 추가적인 손실을 도입하는 것이 도움이 된다는 것을 발견했습니다. 특히, 우리는 Taigman 등의 기술을 채택했습니다. [49] 및 와 같이 대상 도메인의 실제 샘플이 생성기에 입력으로 제공될 때 생성기를 ID 매핑에 가깝게 정규화합니다.

텍스트, 다른, 같은, 다양한이(가) 표시된 사진

자동 생성된 설명

Figure 9: The effect of the identity mapping loss on Monet’s painting→ photos. From left to right: input paintings, CycleGAN without identity mapping loss, CycleGAN with identity mapping loss. The identity mapping loss helps preserve the color of the input paintings.

그림 9: 모네의 회화→사진에 대한 ID 매핑 손실의 영향. 왼쪽에서 오른쪽으로: 입력 그림, ID 매핑 손실이 없는 CycleGAN, ID 매핑 손실이 있는 CycleGAN. ID 매핑 손실은 입력 그림의 색상을 유지하는 데 도움이 됩니다.

Without , the generator G and F are free to change the tint of input images when there is no need to. For example, when learning the mapping between Monet’s paintings and Flickr photographs, the generator often maps paintings of daytime to photographs taken during sunset, because such a mapping may be equally valid under the adversarial loss and cycle consistency loss. The effect of this identity mapping loss are shown in Figure 9.

가 없으면 생성기 G와 F는 필요하지 않을 때 입력 이미지의 색조를 자유롭게 변경할 수 있습니다. 예를 들어, Monet의 그림과 Flickr 사진 간의 매핑을 학습할 때 생성기는 종종 낮의 그림을 일몰 동안 찍은 사진에 매핑합니다. 이러한 매핑은 적대적 손실과 주기 일관성 손실에서 동일하게 유효할 수 있기 때문입니다. 이 ID 매핑 손실의 효과는 그림 9에 나와 있습니다.

In Figure 12, we show additional results translating Monet’s paintings to photographs. This figure and Figure 9 show results on paintings that were included in the training set, whereas for all other experiments in the paper, we only evaluate and show test set results. Because the training set does not include paired data, coming up with a plausible translation for a training set painting is a nontrivial task. Indeed, since Monet is no longer able to create new paintings, generalization to unseen, “test set”, paintings is not a pressing problem

그림 12에서는 Monet의 그림을 사진으로 변환하는 추가 결과를 보여줍니다. 이 그림과 그림 9는 훈련 세트에 포함된 그림에 대한 결과를 보여 주는 반면, 논문의 다른 모든 실험에서는 테스트 세트 결과만 평가하고 보여줍니다. 훈련 세트에는 쌍을 이루는 데이터가 포함되어 있지 않기 때문에 훈련 세트 페인팅에 대한 그럴듯한 번역을 찾는 것은 간단한 작업이 아닙니다. 실제로 모네는 더 이상 새로운 그림, 보이지 않는 일반화, "테스트 세트"를 만들 수 없기 때문에 그림은 시급한 문제가 아닙니다.

Photo enhancement (Figure 14) We show that our method can be used to generate photos with shallower depth of field. We train the model on flower photos downloaded from Flickr. The source domain consists of flower photos taken by smartphones, which usually have deep DoF due to a small aperture. The target contains photos captured by DSLRs with a larger aperture. Our model successfully generates photos with shallower depth of field from the photos taken by smartphones.

사진 향상(그림 14) 우리는 우리의 방법이 더 얕은 피사계 심도로 사진을 생성하는 데 사용할 수 있음을 보여줍니다. Flickr에서 다운로드한 꽃 사진으로 모델을 훈련합니다. 소스 도메인은 스마트폰으로 촬영한 꽃 사진으로 구성되며 일반적으로 조리개가 작아서 DoF가 깊습니다. 대상에는 더 큰 조리개로 DSLR로 캡처한 사진이 포함됩니다. 우리 모델은 스마트폰으로 찍은 사진에서 얕은 피사계 심도의 사진을 성공적으로 생성합니다.

Comparison with Gatys et al. [13] In Figure 15, we compare our results with neural style transfer [13] on photo stylization. For each row, we first use two representative artworks as the style images for [13]. Our method, on the other hand, can produce photos in the style of entire collection. To compare against neural style transfer of an entire collection, we compute the average Gram Matrix across the target domain and use this matrix to transfer the “average style” with Gatys et al [13].

Gatys et al.과의 비교 [13] 그림 15에서는 사진 양식화에 대한 신경 스타일 전달[13]과 결과를 비교합니다. 각 행에 대해 먼저 [13]의 스타일 이미지로 두 개의 대표적인 아트웍을 사용합니다. 반면에 우리의 방법은 전체 컬렉션의 스타일로 사진을 생성할 수 있습니다. 전체 컬렉션의 신경 스타일 전달과 비교하기 위해 대상 도메인에 걸쳐 평균 그람 행렬을 계산하고 이 행렬을 사용하여 Gatys et al[13]과 함께 "평균 스타일"을 전송합니다.

Figure 16 demonstrates similar comparisons for other translation tasks. We observe that Gatys et al. [13] requires finding target style images that closely match the desired output, but still often fails to produce photorealistic results, while our method succeeds to generate natural-looking results, similar to the target domain.

그림 16은 다른 번역 작업에 대한 유사한 비교를 보여줍니다. 우리는 Gatys et al. [13]은 원하는 출력과 거의 일치하는 대상 스타일 이미지를 찾아야 하지만 여전히 사실적인 결과를 생성하는 데 실패하는 반면 우리의 방법은 대상 도메인과 유사한 자연스러운 결과를 생성하는 데 성공합니다.

**6. Limitations and Discussion**

Although our method can achieve compelling results in many cases, the results are far from uniformly positive. Figure 17 shows several typical failure cases. On translation tasks that involve color and texture changes, as many of those reported above, the method often succeeds. We have also explored tasks that require geometric changes, with little success. For example, on the task of dog→cat transfiguration, the learned translation degenerates into making minimal changes to the input (Figure 17). This failure might be caused by our generator architectures which are tailored for good performance on the appearance changes. Handling more varied and extreme transformations, especially geometric changes, is an important problem for future work.

우리의 방법은 많은 경우에 매력적인 결과를 얻을 수 있지만 결과는 균일하게 긍정적인 것과는 거리가 멉니다. 그림 17은 몇 가지 일반적인 실패 사례를 보여줍니다. 위에서 보고된 것과 같이 색상 및 질감 변경을 포함하는 번역 작업에서 이 방법은 종종 성공합니다. 또한 기하학적 변화가 필요한 작업을 탐색했지만 거의 성공하지 못했습니다. 예를 들어, dog→cat 변형 작업에서 학습된 번역은 입력에 대한 최소한의 변경만 수행하도록 퇴보합니다(그림 17). 이 실패는 모양 변경에 대한 우수한 성능을 위해 맞춤화된 생성기 아키텍처로 인해 발생할 수 있습니다. 더 다양하고 극단적인 변형, 특히 기하학적 변화를 처리하는 것은 향후 작업에서 중요한 문제입니다.

Some failure cases are caused by the distribution characteristics of the training datasets. For example, our method has got confused in the horse → zebra example (Figure 17, right), because our model was trained on the wild horse and zebra synsets of ImageNet, which does not contain images of a person riding a horse or zebra.

일부 실패 사례는 훈련 데이터 세트의 분포 특성으로 인해 발생합니다. 예를 들어 말 → 얼룩말 예제(그림 17, 오른쪽)에서 우리의 방법이 혼란스러워졌습니다. 왜냐하면 우리 모델은 말이나 얼룩말을 타는 사람의 이미지가 포함되지 않은 ImageNet의 야생마 및 얼룩말 합성 집합에서 훈련되었기 때문입니다.

We also observe a lingering gap between the results achievable with paired training data and those achieved by our unpaired method. In some cases, this gap may be very hard – or even impossible – to close: for example, our method sometimes permutes the labels for tree and building in the output of the photos→labels task. Resolving this ambiguity may require some form of weak semantic supervision. Integrating weak or semi-supervised data may lead to substantially more powerful translators, still at a fraction of the annotation cost of the fully-supervised systems.

우리는 또한 쌍을 이루는 훈련 데이터로 달성할 수 있는 결과와 쌍을 이루지 않은 방법으로 달성한 결과 사이에 남아 있는 격차를 관찰합니다. 어떤 경우에는 이 간격을 좁히기가 매우 어렵거나 심지어 불가능할 수도 있습니다. 예를 들어, 우리 방법은 때때로 사진 → 레이블 작업의 출력에서 ​​나무와 건물에 대한 레이블을 치환합니다. 이 모호성을 해결하려면 일종의 약한 의미 감독이 필요할 수 있습니다. 약하거나 반 감독된 데이터를 통합하면 여전히 완전 감독 시스템의 주석 비용의 일부만 사용하면서 훨씬 더 강력한 번역기를 만들 수 있습니다.

Nonetheless, in many cases completely unpaired data is plentifully available and should be made use of. This paper pushes the boundaries of what is possible in this “unsupervised” setting.

그럼에도 불구하고 많은 경우에 완전히 페어링되지 않은 데이터를 충분히 사용할 수 있으며 사용해야 합니다. 이 백서는 이 "비감독" 환경에서 가능한 것의 한계를 뛰어 넘습니다.

Acknowledgments: We thank Aaron Hertzmann, Shiry Ginosar, Deepak Pathak, Bryan Russell, Eli Shechtman, Richard Zhang, and Tinghui Zhou for many helpful comments. This work was supported in part by NSF SMA1514512, NSF IIS-1633310, a Google Research Award, Intel Corp, and hardware donations from NVIDIA. JYZ is supported by the Facebook Graduate Fellowship and TP is supported by the Samsung Scholarship. The photographs used for style transfer were taken by AE, mostly in France.

감사의 말: 많은 유용한 의견을 주신 Aaron Hertzmann, Shiry Ginosar, Deepak Pathak, Bryan Russell, Eli Shechtman, Richard Zhang 및 Tinghui Zhou에게 감사드립니다. 이 작업은 NSF SMA1514512, NSF IIS-1633310, Google Research Award, Intel Corp 및 NVIDIA의 하드웨어 기부에 의해 부분적으로 지원되었습니다. JYZ는 Facebook Graduate Fellowship의 지원을 받고 TP는 Samsung Scholarship의 지원을 받습니다. 스타일 트랜스퍼에 사용된 사진은 AE가 주로 프랑스에서 촬영했습니다.

텍스트, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 10: Collection style transfer I: we transfer input images into the artistic styles of Monet, Van Gogh, Cezanne, and Ukiyo-e. Please see our website for additional examples.

그림 10: 컬렉션 스타일 전송 I: 입력 이미지를 Monet, Van Gogh, Cezanne 및 Ukiyo-e의 예술적 스타일로 전송합니다. 추가 예제는 당사 웹사이트를 참조하십시오.

다른이(가) 표시된 사진

자동 생성된 설명

Figure 11: Collection style transfer II: we transfer input images into the artistic styles of Monet, Van Gogh, Cezanne, Ukiyo-e. Please see our website for additional examples.

그림 11: 컬렉션 스타일 전송 II: 입력 이미지를 Monet, Van Gogh, Cezanne, Ukiyo-e의 예술적 스타일로 전송합니다. 추가 예제는 당사 웹사이트를 참조하십시오.

텍스트, 다른, 다양한이(가) 표시된 사진

자동 생성된 설명

Figure 12: Relatively successful results on mapping Monet’s paintings to a photographic style. Please see our website for additional examples.

그림 12: Monet의 그림을 사진 스타일로 매핑한 비교적 성공적인 결과. 추가 예제는 당사 웹사이트를 참조하십시오.

다른, 다양한, 여러개이(가) 표시된 사진

자동 생성된 설명

Figure 13: Our method applied to several translation problems. These images are selected as relatively successful results – please see our website for more comprehensive and random results. In the top two rows, we show results on object transfiguration between horses and zebras, trained on 939 images from the wild horse class and 1177 images from the zebra class in Imagenet [5]. Also check out the horse→zebra demo video. The middle two rows show results on season transfer, trained on winter and summer photos of Yosemite from Flickr. In the bottom two rows, we train our method on 996 apple images and 1020 navel orange images from ImageNet.

그림 13: 우리의 방법은 여러 번역 문제에 적용되었습니다. 이 이미지는 비교적 성공적인 결과로 선택되었습니다. 보다 포괄적이고 무작위적인 결과를 보려면 당사 웹사이트를 참조하십시오. 맨 위 두 행에는 Imagenet[5]에서 야생마 클래스의 939개 이미지와 zebra 클래스의 1177개 이미지에 대해 훈련된 말과 얼룩말 간의 객체 변형에 대한 결과가 표시됩니다. 말→얼룩말 데모 영상도 확인하세요. 가운데 두 줄은 Flickr의 요세미티 겨울 및 여름 사진으로 훈련된 시즌 이적에 대한 결과를 보여줍니다. 맨 아래 두 행에서 우리는 ImageNet의 996개의 사과 이미지와 1020개의 배꼽 오렌지 이미지에 대해 방법을 훈련합니다.

꽃, 식물, 실외, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 14: Photo enhancement: mapping from a set of smartphone snaps to professional DSLR photographs, the system often learns to produce shallow focus. Here we show some of the most successful results in our test set – average performance is considerably worse. Please see our website for more comprehensive and random examples.

그림 14: 사진 향상: 스마트폰 스냅 세트에서 전문 DSLR 사진으로 매핑하면 시스템이 종종 얕은 초점을 생성하는 방법을 학습합니다. 여기에서는 테스트 세트에서 가장 성공적인 결과 중 일부를 보여줍니다. 평균 성능은 상당히 나쁩니다. 보다 포괄적이고 무작위적인 예는 당사 웹사이트를 참조하십시오.

텍스트, 다른, 다양한, 다채로운이(가) 표시된 사진

자동 생성된 설명

Figure 15: We compare our method with neural style transfer [13] on photo stylization. Left to right: input image, results from Gatys et al. [13] using two different representative artworks as style images, results from Gatys et al. [13] using the entire collection of the artist, and CycleGAN (ours).

그림 15: 사진 양식화에 대한 신경 스타일 전송[13]과 방법을 비교합니다. 왼쪽에서 오른쪽으로: 입력 이미지, Gatys et al. [13] 두 개의 서로 다른 대표 작품을 스타일 이미지로 사용하여 Gatys et al. [13] 아티스트의 전체 컬렉션과 CycleGAN(당사)을 사용합니다.

텍스트, 다른, 묶음, 다양한이(가) 표시된 사진

자동 생성된 설명

Figure 16: We compare our method with neural style transfer [13] on various applications. From top to bottom: apple→orange, horse→zebra, and Monet→photo. Left to right: input image, results from Gatys et al. [13] using two different images as style images, results from Gatys et al. [13] using all the images from the target domain, and CycleGAN (ours).

그림 16: 다양한 애플리케이션에서 우리의 방법을 신경 스타일 전달[13]과 비교합니다. 위에서 아래로 사과→오렌지, 말→얼룩말, 모네→사진. 왼쪽에서 오른쪽으로: 입력 이미지, Gatys et al. [13] 두 개의 다른 이미지를 스타일 이미지로 사용하여 Gatys et al. [13] 대상 도메인의 모든 이미지와 CycleGAN(당사)을 사용합니다.

텍스트, 다른이(가) 표시된 사진

자동 생성된 설명

Figure 17: Typical failure cases of our method. Left: in the task of dog→cat transfiguration, CycleGAN can only make minimal changes to the input. Right: CycleGAN also fails in this horse → zebra example as our model has not seen images of horseback riding during training. Please see our website for more comprehensive results.

그림 17: 우리 방법의 일반적인 실패 사례. 왼쪽: 개→고양이 변형 작업에서 CycleGAN은 입력을 최소한으로만 변경할 수 있습니다. 오른쪽: 우리 모델이 훈련 중 승마 이미지를 보지 못했기 때문에 이 말 → 얼룩말 예제에서도 CycleGAN이 실패합니다. 보다 종합적인 결과는 당사 웹사이트를 참조하십시오.