

공간 적응적 비정규화 방식 기반 단일 영상 내 그림자 제거 기법

류현정*, 최윤식°

Spatially Adaptive De-Normalization Based Shadow Removal from a Single Image

Hyunjeong Ryu*, Yoonsik Choe°

요약

단일 영상 내 그림자 제거는 영상 처리 및 많은 컴퓨터 시각 분야에서 중요한 문제이다. 이 논문에서는 이러한 그림자를 제거하기 위하여, 추가적인 그림자 감지를 필요로 하지 않는 조건부 적대적 생성 신경망 기반의 새로운 신경망 구조를 제안한다. 제안하는 구조에서는, 신경망 내 다양하게 존재하는 정규화층으로 인해 발생하는, 입력 영상 데이터의 정보 손실 문제를 방지하기 위해 공간 적응적 비정규화 방식을 활용하였고, 그림자 제거와 관련된 공인 데이터 세트를 활용한 다양한 실험들로부터, 제안된 기법이 기존의 신경망 기반의 최신 방법론들과 비교하여 적어도 5dB 높은 PSNR 성능을 가지는 것을 확인할 수 있다.

키워드 : 조건부 적대적 생성 신경망, 정규화 층, 그림자 감지, 공간 적응적 비 정규화, 그림자 제거

Key Words : Conditional generative adversarial networks, Normalization layers, Shadow detection, Spatially adaptive de-normalization, Shadow removal

ABSTRACT

Shadow removal from a single-image has been a significant issue in image processing and many computer vision areas. This paper proposes a novel network based on the conditional generative adversarial network scheme without requiring additional shadow detection process to remove shadows. The proposed structure also utilizes a spatially adaptive de-normalization method to prevent the input image information loss caused by various normalization layers in the neural network. From the various experiments related to shadow removal using authorized datasets, it is confirmed that the proposed network shows at least 5dB higher performance in PSNR, compared to the state of the arts neural network based methodologies.

1. 서론

그림자란 그림 1에 보이는 것과 같이 물체 뒤쪽으로 빛이 통과하지 못하면서 생기는 어두운 부분을 말

한다. 이러한 영상 내 그림자는, 단일 영상으로부터, 영상 분할이나 객체 인식 등과 같은 다양한 영상 처리 및 컴퓨터 시각 처리를 할 때, 마치 별도의 물체가 존재하는 것처럼 작용하여 원하는 객체만을 검출해내

* 본 연구는 문화체육관광부 및 한국 콘텐츠진흥원의 연구 개발지원사업으로 수행되었음(과제번호: R2020040239)

• First Author : Yonsei University Department of Electrical & Electronic Engineering, mesh0213@yonsei.ac.kr, 정회원

° Corresponding Author : Yonsei University Department of Electrical & Electronic Engineering, yschoe@yonsei.ac.kr, 종신회원
논문번호 : 202107-157-A-RN, Received July 7, 2021; Revised September 19, 2021; Accepted January 12, 2022

는 것을 매우 어렵게 만든다.^[1,2] 또한, 그림자는 광원의 종류에 따라 다양한 형태로 나타날 수 있기 때문에 단일 영상 내에서 그림자 영역만을 제거해낸다는 것은 쉽지 않다. 하지만 최근에는 심층 학습기법을 기반으로 한 영상 인식 기술의 발전과 더불어 그림자 제거 또한 다양한 연구^[3-10]가 이루어졌다.

기존의 적대적 생성 신경망 (generative adversarial network)^[6]만을 이용하여 그림자 제거를 하게 되면 원하는 방향의 영상을 얻기 어렵기 때문에 그림자 연구의 대부분은 원하는 방향의 영상을 생성해내기 위해 조건을 정해주는 조건부 적대적 생성 신경망 (conditional generative adversarial network)^[7]을 기반으로 진행되어왔다. 하지만 이러한 기존의 연구 결과에서도 잡음이나 잔상 형태로 제거되지 않은 영역이 그림자와 객체의 경계면에 남아있게 될 뿐만 아니라 그림자 영역과 주변 영역과의 색상 불일치 문제 역시 나타난다. 이러한 문제점들을 해결하기 위해서 최근에는 그림자 감지를 위한 신경망과 그림자 제거를 위한 신경망 둘을 동시에 학습하는 방법이 제안되었다.^[8] 그러나 이 방법 또한 그림자 영역에만 집중하는 것이 아니라 영상 내에 다양한 배경과 물체의 정보를 함께 고려하기 때문에 여전히 그림자 영역을 제대로 제거하지 못하고 배경색상 역시 복원하지 못하는 문제가 있다.

이 논문은 최근 그림자 제거 연구 동향처럼 심층 신경망을 활용하여 영상데이터 간 전환문제^[9]를 기반으로 한 새로운 신경망을 제안한다. 영상 데이터 간 전환은 별도의 그림자 영역에 대한 정보 없이 그림자가 있는 영상 데이터와 그림자가 없는 영상 데이터를 각각 서로 데이터 군집으로 정의하고, 그림자 영역이 포함된 영상으로부터 그림자가 제거된 영상 데이터로의 변환을 가능케 하는 함수를 심층 신경망을 통해 추정하는 것이다. 이를 위해 조건부 적대적 생성신경망을 활용하여 원하는 방향의 영상을 얻을 수 있도록 한다. 다수의 기존 조건부 적대적생성신경망에서는 생성기

의 학습 수렴을 빠르게 하기 위해 배치 정규화,^[10,11] 순간 정규화^[12]등 다양한 정규화 층을 활용하여 수렴을 유도하였다. 하지만, 그림자 제거와 같이 성능이 입력 영상의 다양한 공간적 특징에 매우 민감한 경우, 다양한 정규화 과정으로 인해 오히려 입력 영상 내 다량의 정보가 손실되어 성능이 제한되는 문제점이 발생된다.

이 논문에서는 이 문제를 추가적으로 해결하기 위해 공간 적응적 비 정규화 (spatially adaptive denormalization) 방식을 활용한다.^[13] 공간 적응적 비 정규화 방식은 복호기에 입력 영상의 공간 정보를 담아 요소별 연산을 통해 결합해줌으로써 생성기 내의 복호기에서 입력 영상의 다양한 공간적 특징을 활용할 수 있도록 한다. 또한, 그림자의 경계면과 객체의 경계면을 잘 구분해낼 수 있도록 다양한 해상도의 형태로 그림자 영상의 정보를 넣어줌으로써 넓은 범위에서 좁은 범위까지의 다양한 특징 모두를 담은 특성 맵을 통해 입력 영상의 해상도에 따른 계층적 특징 또한 보장할 수 있게 된다. 이 때문에 그림자와 객체의 경계면을 잘 구분할 뿐만 아니라 입력 영상의 정보를 손실하지 않고 잘 보존할 수 있게 된다.

이 논문의 구성은 다음과 같다. II장에서는 제안된 신경망의 구조를 포함하여 제안하는 방법론을 구체화하여 나타내고, III장에서는 기존의 방법론과의 비교 평가를 통해 정성 및 정량적인 수치로 성능을 나타낸다. 마지막으로 IV장에서는 제안하는 방법론의 최종 결론을 기술하였다.

II. 본 론

본 장에서는 단일 영상 내 그림자 영역 제거 문제를 영상 데이터 간 전환 문제로 정의하여 조건부 적대적 생성신경망을 활용한 기존 방법론에 대하여 서술한 뒤 논문에서 제안하는 기법을 설명한다.

2.1 단일 영상 내 그림자 분해를 통한 그림자 제거 기법

초기의 그림자 제거 연구는 물리적 모형을 기반으로, 그림자 없는 영상, 그림자 있는 영상, 재조명된 영상, 그리고 그림자 영역을 담은 마스크 정보로 구성하였고, 문제 정의는 그림 2에 나타난 것과 같다. 즉, 그림자 없는 영상은 아래 식(1)과 식(2)에 의하여 표현되고, 식에 포함된 계수 α , ω , b 를 예측함으로써, 그림자 있는 영상으로부터 분리할 수 있는 방법이 제안되었다.^[14]



그림 1. 실제 그림자의 예
Fig. 1. shadow examples in real world

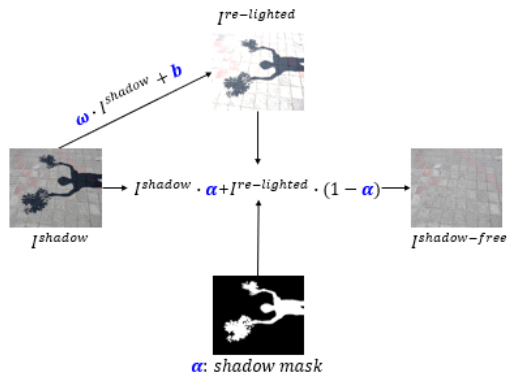


그림 2. 그림자 분해 기법을 통한 그림자 제거 방법론[14]
Fig. 2. shadow removal via shadow decomposition

$$I^{shadow-free} = I^{shadow} \cdot \alpha + I^{re-lighted} \cdot (1 - \alpha) \quad (1)$$

위 식에서 I^{shadow} 는 그림자 영상, $I^{shadow-free}$ 는 그림자 없는 영상을 의미한다. 매개변수 α 는 그림자 마스크의 화소 단위 정보를 나타내며, 이상적으로는 비그림자 영역은 1, 그림자 영역은 0의 값을 가진다. 따라서, 식(1)은 그림자 없는 영상은 재 조명된 영상과 그림자 영상의 화소 단위의 선형결합임을 나타낸다. 또한 재 조명된 영상 $I^{re-lighted}$ 의 픽셀의 정보는 아래 수식(2)와 같이 연산된다.

$$I^{re-lighted} = \omega \cdot I^{shadow} + b \quad (2)$$

여기서 $I^{re-lighted}$ 는 재 조명된 영상을 나타내며, ω 와 b 는 그림자가 있는 영상의 픽셀에서 재 조명된 영

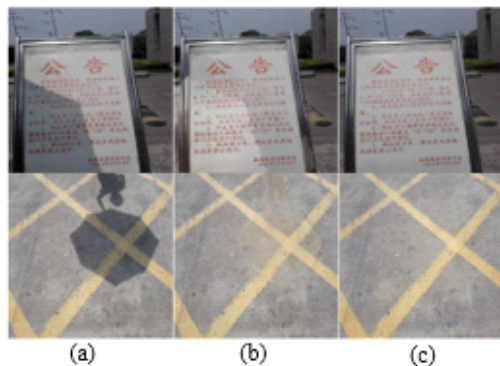


그림 3. 기존 물리적 모형을 기반으로 한 그림자 분리 방법의 실패 사례
Fig. 3. Failure cases of prior physical modeling based on shadow removal method

상의 화소단위 변환 시 사용되는 매개변수이다.

하지만, 위와 같이 그림자 영상 I^{shadow} 로부터 그림자 마스크를 이용하여 그림자 없는 영상 $I^{shadow-free}$ 를 분리해내는 작업을 하더라도, 예측 오차에 의하여 그림자와 객체 간의 경계면에 그림 3과 같이 잡음 형태로 그림자의 형태가 남아있는 현상이 나타난다. 그림 3에서 (a)는 그림자가 있는 입력 영상, (b)는 물리적 모형을 기반으로 연구된 SP+M-Net 모형[9]의 실험 결과 영상, (c)는 정답 영상을 각각 나타낸 그림이다. 이처럼 그림자 영역에 대한 마스크 정보가 있더라도 그림자의 경계면과 물체의 경계면을 제대로 구분해 내지 못하는 문제가 존재한다.

2.2 심층 학습 기반의 그림자 제거 방법론

앞 절에서 다룬 바와 같이 그림자 제거 문제는 매개변수로부터 영상을 생성해내는 것이 아닌, 그림 4와 같이 그림자 영상 A에서 그림자 없는 영상 B로 변환하는 것이기 때문에 이는 영상 데이터 간 전환으로 생각할 수 있다.

영상 데이터 간 전환 방법 가운데 가장 대표적인 방법론은 PIX2PIX[9]로, 이는 조건부 적대적 생성 신경망의 가장 대표적인 방법이다. 그림 5는 이 신경망 구조를 활용한 그림자 제거 방법론을 이해하기 쉽게 나타낸 것이다. 이 신경망의 경우는 기존의 확률잡음 z 로부터 출력벡터 y 를 생성하는 적대적 생성 신경망과는 달리, 잡음 벡터 z 뿐만 아니라, 입력 벡터 x 를 활용하여 출력벡터 y 를 생성하는 조건부 적대적 생성 신경망을 활용한다. 생성기 G 는 특정 조건을 받아 무엇을 출력할 것인가에 대한 방향성을 가지게 되고, 판별기 D 도 특정 조건을 받아 무엇이 올 것인지 기대하며 예측을 하게 된다. 조건부 적대적 생성 신경망의 목적 함수는 다음 수식 (3)과 같이 정의되고, 수식 (4)의 경우는 조건부가 아닌 기존 적대적 생성 신경망의

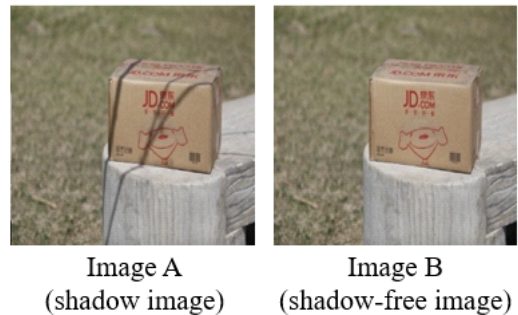


그림 4. 영상 A에서 영상 B로의 전환 방법론
Fig. 4. Translation Image A to Image B

목적 함수이다. 여기서 x, y 는 각각 입력벡터와 출력 벡터를 나타내며, z 는 잡음벡터를 의미하고, E 는 기댓값을 뜻한다.

$$L_{CGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z)))] \quad (3)$$

$$L_{GAN}(G, D) = E_{x,y}[\log D(y)] + E_{x,z}[\log(1 - D(G(x, z)))] \quad (4)$$

식 (3)과 (4)의 차이는, 판별기 D 에 들어가는 입력의 형태이다. 즉, 식 (3)은 입력 영상 데이터를 특정 조건으로 활용하여 해당 조건을 만족시키는 영상 데이터 생성을 가능케 하는 방식이다. 이와 같은 차이점을 통해 조건부 적대적 생성신경망의 판별기 D 는 입력 영상과 출력 영상이 서로 올바른 짝을 이루는지 판별하고, 생성기는 입력받은 x 에 올바른 짝을 이루는 y 와 유사한 영상 데이터를 생성해내기 위해 학습이 진행되는 구조를 가진다.^[15,16]

2.3 공간 적응적 비정규화 기법을 기반으로 한 단일 영상 내 그림자 제거 방법론

앞 절에서 설명한 것과 같이 기존의 적대적 생성 신경망을 활용한 구조에서는 생성기의 학습 수렴을 가속화 하기 위해 다수의 배치 정규화 혹은 순간 정규화와 같은 다양한 정규화를 활용하였다. 그러나, 이러한 다수의 정규화 층은, 신경망이 안정적인 수렴을 보장하는 특성이 있음에도 불구하고, 그림자 영상과 같이 생성기가 복잡한 물리적 현상을 오직 입력 영상만을 활용하여 추론하는 경우, 영상 데이터 정보의 손실 문제를 피할 수 없게 된다. 영상 데이터 정보의 손실

문제란, 다수의 정규화 층을 통과하면서 입력 영상 내 존재하는 다량의 중요한 특성-맵 정보들이 손실되는 현상을 말한다. 이러한 문제를 해결하기 위해 생성기를 구성하는 복호기에 부호기의 특성-맵을 결합하는 형태의 신경망 구조가 제안되었지만^[17], 이 경우 부호기가 입력 영상 데이터를 잠재적인 특징 공간에 투사하는 것만을 목적으로 하기 때문에, 부호기는 복호기의 특성-맵에 단순 결합하는 구조를 가지고, 이는 단순히 입력 영상의 공간적 특징의 일부 반영은 가능하나, 근본적인 해결 방법은 되지 못한다.

따라서, 이 논문에서는 기존의 조건부 적대적 생성 신경망을 기반으로 공간 적응적 비정규화 기법^[13]을 활용하여 생성기의 수렴은 보장하고, 입력 영상 데이터의 정보 손실을 원칙적으로 차단하도록 구조를 제안하였다.

생성기 G 를 구성하는 특성-맵을 F 라고 하고, 입력 그림자 영상을 S 라고 하면, 공간 적응적 비 정규화 기법은 아래 수식 (5)과 같이 표기된다.

$$SPADE(F, S) = \gamma(S) \otimes \frac{F - \mu_F}{\sigma_F} \oplus \beta(S) \quad (5)$$

수식 (5)에서, μ_F 와 σ_F 는 공간 적응적 비정규화가 적용될 특성-맵 F 를 구성하는 요소들의 평균과 표준편차를 의미하며, \otimes 와 \oplus 는 요소별 곱 과 요소별 합을 나타낸다. 그러므로 계수로 사용된 $\gamma(S)$ 와 $\beta(S)$ 는 학습된 값이 아닌 중첩층에서 얻어지므로, 특성-맵과 같은 특성을 가진다. 이 때, 그림자 영상의 특정 공간에 대한 정보는 위의 γ 와 β 의 형태로 삼입되므로, 공간적인 특징 정보가 보존되어 기존의 배치 정규화

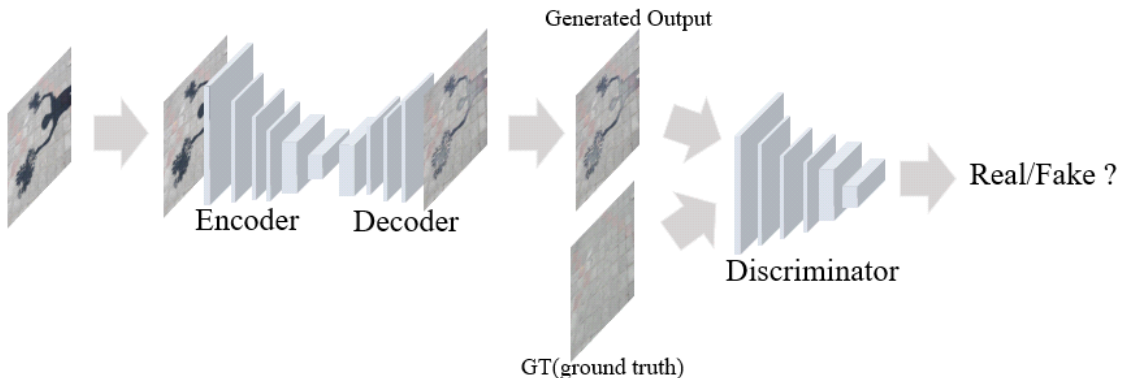


그림 5. PIX2PIX 신경망 구조
Fig. 5. PIX2PIX network structure

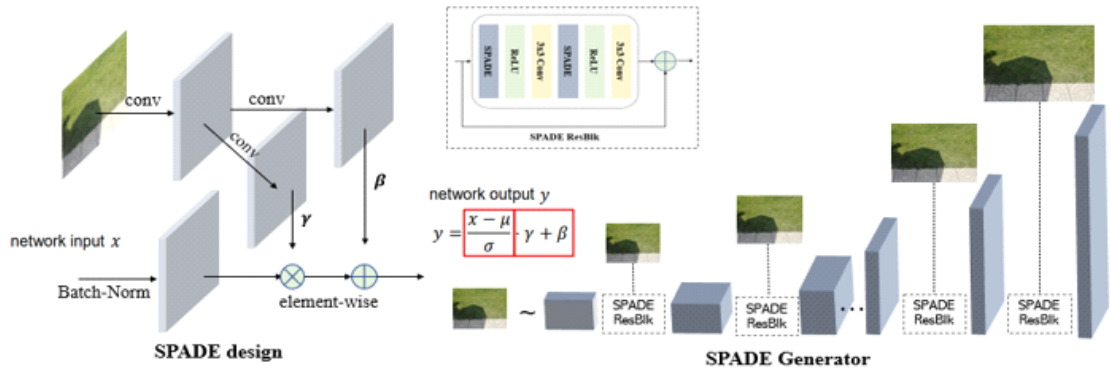


그림 6. (좌)공간 적응적 비 정규화 기법 (우) 공간 적응적 비 정규화 기법 생성기의 다양한 해상도의 계층적 특징을 보존한 구조 Fig. 6. (left)spatially adaptive de-normalization structure, (right)multi-resolution structure of spatially adaptive de-normalization generator

가 가지고 있던 영상 데이터 정보 손실의 문제가 해결된다. 수식 (5)에서 표현된 전체 과정은 그림 6을 통해 확인할 수 있다.

이 논문에서는 다양한 해상도 구조에 삽입되는 공간 적응적 비 정규화 기법을 활용하여 생성기를 구성한다. 이는 다양한 특징 정보들의 손실을 방지할 뿐만 아니라, 생성기의 복호기를 구성하는 다양한 크기의 특성-맵에 적용되어 입력 영상 데이터의 크기에 따른 계층적 특징 또한 보장할 수 있게 해 준다. 또한, 그림 6의 우측과 같은 구조의 다수의 계층적 특징 정보를 삽입하는 최적 층의 수는 III장에서 성능의 차이 분석을 위한 실험을 통하여 결정하였다.

2.4 제안하는 신경망 구조

본 절에서는, 그림 7의 신경망 구조와 같이, 특정 조건을 받아 무엇을 출력할 것인지에 대한 방향성을 가지는 조건부 적대적생성신경망 구조를 기반으로, 단순히 PIX2PIX만을 사용한 신경망에서 발생할 수 있는 입력 영상 정보의 손실을 막을 수 있도록 공간 적응적 비 정규화 기법을 활용한 새로운 신경망을 제안한다. 구조적으로 이 신경망은 별도의 그림자 영역을 가진 마스크 검출이나, 그림자 영역을 감지하는 모듈을 필요로 하지 않도록 구성되었다. 제안된 신경망은 그림 5와 같이 생성기와 판별기로 이루어진 적대적생성신경망 구조를 기반으로 하되, 입력 영상에 공간 적응적 비정규화 방식을 활용하여 그림자 영상을 복호

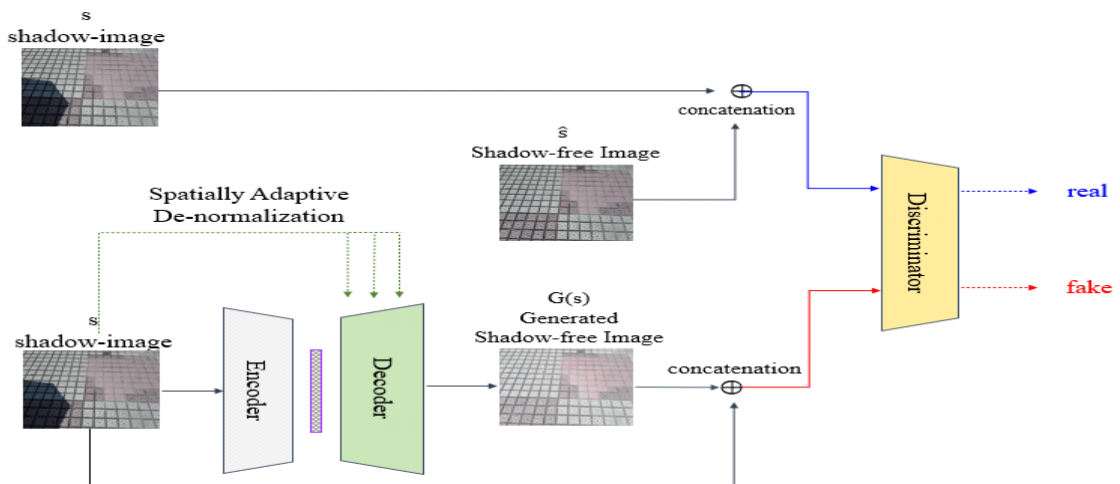


그림 7. 제안하는 알고리즘의 전체 구조: 별도의 그림자 마스크 정보 없이 조건부 적대적 생성신경망을 기반으로 공간 적응적 비 정규 기법을 활용한 단일 영상 내 그림자 제거 기법 Fig. 7. Proposed network, A method of removing shadow regions within a single image using a spatially adaptive de-normalization based conditional adversarial generative neural network without separate shadow mask information

기에 삽입해줌으로써 공간적인 특징을 보존할 수 있도록 설계되었다.

그림 7과 같이, 입력된 그림자 영상 s 를 입력으로 받아 공간 적응적 비 정규화로 구성된 생성기를 통과하여 얻어진 영상 $G(s)$ 는, 그림자 영상 s 와 한 쌍을 이루게 되고, 그림자 없는 영상 \hat{s} 는 그림자 영상 s 와 또 다른 한 쌍을 이루게 되어, 이 두 쌍을 판별기가 입력으로 받아 진짜 그림자 없는 영상인지 생성기로부터 만들어진 가짜 그림자 없는 영상인지를 판단하게 되는 구조로 구성된다. 이렇게 판별기가 두 영상 $((s, G(s)), (s, \hat{s}))$ 을 구분해내는 목적 함수는 다음의 수식 (6)와 같이 영상들 차이의 기댓값 E 들의 합으로 정의된다.

$$\min_G \max_D V(D, G) = E_{s, \hat{s}} [\log D(s, \hat{s})] + E_s [1 - \log D(s, G(s))] \quad (6)$$

생성기 G 는 판별기 D 의 오류값을 감소시키도록 학습이 될 뿐만 아니라 $G(s)$ 가 그림자 없는 영상 \hat{s} 와 같아지도록 학습하게 된다. 이처럼, 잘 학습된 결과는 생성기 G 로부터 최종적으로 그림자 없는 영상을 생성하게 된다.

결론적으로, 제안된 신경망 구조는, 원하는 영상의 방향성을 더해주는 조건부 적대적 생성 신경망을 기반으로 하되, 별도의 그림자 마스크 정보를 활용하지 않고도 영상의 공간적인 주요 특징들이 손실되지 않도록 한, 공간 적응적 비정규화 기법을 활용한 신경망 구조이다.

다음 III장에서는 이 논문에서 제안한 신경망의 성능을 실험을 통해서 정성 및 정량적으로 검증한다.

III. 실험

3.1 실험 환경 및 구현 세부 정보

본 장에서는, 실험을 위해 적대적생성신경망을 기반으로 한 기존의 방법론과 그림자 제거 연구의 최신 방법론을 선정하여 비교실험을 진행하였다. 적대적 생성 신경망 기반의 가장 대표적인 PIX2PIX방식⁹⁾과, 이를 기반으로 그림자 제거와 그림자 감지를 동시에 학습하면서, 상호 간의 정보 흐름이 공유되도록 제안된 ST-CGAN⁸⁾ 방식을 추가적으로 비교하였다. 뿐만 아니라, 그림자 마스크 정보 없이는 그림자 없는 영상을 얻을 수 없는 물리적 특성을 활용한 SP+M-Net방식¹⁴⁾을 비교 방식으로 선정하여, 이 논문에서 제안한

신경망이 별도의 그림자 마스크 정보 없이도 얼마나 좋은 성능을 가지는지를 검증하였다. 또한 최근의 그림자 제거 연구인 DHAN방식¹⁸⁾도 비교실험 모형으로 활용하였는데, 이는 그림자 마스크를 활용하여 그림자 영상을 합성하는 그림자 실험 데이터를 증가시키고, 제한된 데이터 집합 학습에 의한 그림자 경계면 잡음이나 잔상이 없는 고품질 영상을 생성하도록 제안된 신경망이다. 이 연구 또한 그림자 마스크를 활용한 데이터 집합의 확장을 통해 그림자 영역에 대한 학습을 강화하여 성능을 높이는 최신 연구이기에 이 실험에서도 비교 모형으로 활용한다. 제안된 방법과 비교실험을 위한 방법들의 구현은 파이토치를 활용하였고, 대표적인 데이터 집합인 ISTD데이터 집합⁸⁾과 SRD데이터 집합¹¹⁾을 활용하여 실험을 진행하였다. ISTD데이터 집합은 총 1840쌍의 영상 데이터들로 구성되어 있으며, 학습용 데이터 1300쌍, 나머지 540쌍의 영상 데이터들을 시험용 데이터 집합으로 구분한다. SRD 데이터 집합의 경우는 총 3058쌍으로 구성되며, 학습용 데이터 2650쌍, 나머지 408개의 시험용 데이터 집합으로 구분되었다. 이 두 데이터 구성의 차이점은, ISTD 데이터 집합의 경우 {그림자 영상, 그림자 없는 영상, 그림자 마스크 영상}의 형태로 구성되었고, 반면 SRD 데이터 집합은 {그림자 영상, 그림자 없는 영상}과 같이 두 쌍의 형태로 존재하게 되어, 그림자 마스크 정보가 별도로 존재하지 않는 데이터 집합의 경우에 대하여도 성능 분석이 가능하다. 따라서, SRD 데이터 집합은 기존 방식 가운데 그림자 마스크 정보를 사용하지 않는 경우와의 비교 실험에 활용하였다. 실험에서 사용된 평가 방법은 PSNR(Peak Signal-to-Noise Ratio)과 RMSE(Root mean squared error)이며, 그림자 없는 영상과의 비교를 통해 정성적인 결과 뿐만 아니라, 정량적인 수치를 통해 이 논문에서 제안한 신경망의 우수한 성능을 확인하였다.

3.2 다양한 데이터 집합을 활용한 그림자 제거 결과

먼저, 그림자 마스크 정보를 필요로 하지 않는 기존의 연구 방식 중 하나인 PIX2PIX와 제안된 방법론을 비교하여 실험을 진행하였다. PIX2PIX의 경우는 그림자 영상을 입력으로 받아 출력 영상을 얻는 부호기-복호기 구조로 구성되어, 입력 영상과 출력 영상 간의 일치를 학습할 뿐만 아니라, 이러한 일치를 학습하기 위한 손실함수도 학습을 하게 되는 구조로 구성되어 있다. 하지만, 앞서 언급한 것과 같이 기존의

PIX2PIX 방식은 생성기 G 에 포함된 다수의 배치 정규화 층들이 입력 영상의 공간적인 특징 정보를 손상 시키게 되는 현상으로 인해, 그림자를 제대로 제거하지 못하는 결과를 초래한다. 이를 해결하기 위해 제안된 신경망의 성능은 그림자가 있는 영상 가운데 세부 정보가 많이 포함된 그림자 영상 위주로 실험 결과를 확인하여, 영상의 공간적 특징 정보가 잘 보존되어 그림자가 잘 제거 되는 지를 확인함으로써 성능을 증명하였다. 아래 그림 8은 그림자 영상에 세부 정보가 많은 경우, 제안된 방법이 그림자에 가려진 세부 영역을 잘 표현해 내는 것을 시각적으로 잘 보여준다. 이에 반해 기존의 PIX2PIX방식의 경우는 입력 영상의 정보가 손실되는 문제로 인해 그림자가 지워진 부분이 제대로 표현되지 않음을 확인할 수 있다. 이를 통해 다수의 정규화 층으로 인해 발생하는 입력영상 정보의 손실 문제가 제안된 방법론으로 해결됨을 확인하였다.



그림 8. PIX2PIX방법론과의 그림자 제거 성능 비교
Fig. 8. Results of removing shadow images comparing PIX2PIX method



그림 9. 제안된 방법론의 실패 사례; 왼쪽부터 차례대로 입력 그림자 영상, 제안한 방법론, PIX2PIX, 정답영상(Ground truth)을 나타냄
Fig. 9. Failure cases of the proposed method; It represents input shadow images, proposed methodologies, PIX2PIX, Ground-truth images sequentially from the left

다음은 제안된 기법이 어떠한 경우에 성능의 제한이 발생하는 지에 대한 실험 결과이다. 그림 9은 그림자가 같은 검은색 계열의 객체와 겹쳐져 있는 경우, 객체와 그림자 영역을 제대로 구분해내지 못하는 경우이다. 이는 그림자와 객체의 경계면이 객체의 경계면인지, 그림자의 경계면인지를 제대로 판단하지 못하여, 제안된 기법의 결과에서 조차 그림자 제거가 제대로 이루어지지 않은 경우이다. 이러한 경우는, 경계면에 대한 정보를 더 잘 보존할 수 있도록 더 깊은 계층

적 해상도 구조를 설계함으로써 해결될 수 있고, 결과적으로 그림자와 객체의 경계면에 대한 정보를 더욱 세밀하게 분리 해 낼 수 있다. 이를 확인하기 위하여 추가적으로 수행한 실험의 결과는 표 1과 같다. 즉, 공간 적응적 비정규화를 활용한 생성기에 삽입하는 계층적 입력 영상의 수를 증가시키면서 성능의 변화를 확인한 결과, 공간 적응적 비정규화의 생성기에 삽입되는 입력 영상의 수가 늘어날수록 성능이 좋아지는 것을 확인할 수 있다. 하지만, 층별로 넣어주는 영상

표 1. 공간 적응적 비 정규화 생성기에 넣어주는 입력 영상 정보의 수에 따른 성능 비교 결과; 왼쪽 열부터 차례대로 입력그림자 영상, ground-truth영상, 생성기에 넣어주는 입력 영상 정보의 수(5, 6, 7)을 나타냄. 각각의 행은 서로 다른 그림자 영상에 대한 결과를 나타냄
Table 1. Performance comparison results according to the number of image information put into the spatially adaptive de-normalization generator; It represents the number (5, 6, 7) of input shadow images, ground-truth images, and input image information placed in the generator sequentially from the left column. Each row represents a result for a different shadow image

result	input shadow image	Ground-truth	# of Layers		
			5	6	7
Test 1					
Test 2					

표 2. SRD 데이터집합을 활용한 PIX2PIX 방법론과의 그림자 제거 성능 비교 결과
 Table 2. Results of removing shadow comparison with PIX2PIX method on SRD dataset

(a) shadow image	(b) shadow-free image (Ground-Truth)	(c) ours	(d) PIX2PIX

데이터의 정보가 많아지면 계산 시간이 그만큼 증가하게 되고, 신경망 자체가 복잡해지는 것은 불가피한 한계점이기 때문에 적절한 층수를 선택하는 것은 중요하다. 이 논문에서는 전반적인 성능은 보장하면서 계산 복잡도에 큰 영향이 미치지 않도록 6개의 층을 실험적으로 선정하였다. 결론적으로 논문에서 제안하

는 방법론은, 그림자 마스크 정보를 필요로 하지 않지만, 층마다 입력 영상의 공간적인 특징 정보를 삽입해주는 구조를 채택하여, 다양한 해상도 구조의 유무가 성능의 차이를 나타냄을 확인하였고, 표 2를 통하여서는 다양한 데이터집합으로부터의 결과를 비교 확인하였다. 또한 별도의 그림자 마스크 정보가 필요한 기존

표 3. ISTD 데이터 집합을 활용한 기존 방법론과의 그림자 제거 성능 비교 결과
 Table 3. Results of shadow removing comparison with prior method on ISTD datasets

(a) shadow image. shadow-free image	(b) ours	(c) PIX2PIX	(d) ST-CGAN	(e) Le(SP+M-Net)	(f) DHAN

표 4. ISTD 데이터집합에 대한 PSNR, RMSE 결과
Table 4. PSNR and RMSE results on ISTD Datasets

	Average of PSNR [dB]	Average of RMSE
Ours	30.6	10.2
PIX2PIX	24.7	14.4
ST-CGAN	20.5	18.2
SP+M-Net	21.2	24.7
DHAN	21.1	24.1

의 방법을 함께 포함하여 그림자 마스크 영상을 포함하고 있는 ISTD 데이터 집합에 대해서도 비교한 결과를 표 3에서 확인할 수 있다. 결론적으로, 제안된 신경망 구조는 별도의 그림자 마스크를 감지하거나 활용하지 않고도 기존의 방법론들보다 간단하고, 효율적으로 구성된 신경망 구조를 통해서 우수한 성능의 결과를 얻을 수 있음이 확인되었다. 또한 정량적 평가를 위하여, 그림자 없는 영상과의 상대적 평가 지표인 PSNR과 RMSE 값으로 나타내면 표 4와 같다. 결과적으로 이는, 논문에서 제안한, 별도의 추가과정이 필요없는 독립적인 신경망이, 정성적으로나 정량적으로도 우수한 그림자 제거 성능을 보여줌을 나타낸다.

IV. 결 론

이 논문은 조건부 적대적 심층신경망을 활용한 새로운 그림자 제거 방법을 제안한다. 그림자 제거 문제는 영상 데이터 간의 전환 문제로 정의하였고, 조건부 적대적 심층신경망의 구조를 기반으로 해결하였다. 제안된 기법은, 기존의 다수의 정규화 층으로 인해 발생되었던 입력 영상 데이터 정보의 손실 문제를 해결하였고, 복호화 과정에서 다양한 해상도에 대한 계층적 특징을 활용하여 비슷한 색깔의 물체 위에 형성된 복잡한 그림자 역시도 제거할 수 있었다. 특히 제안된 신경망은 그림자 영역을 별도 감지하는 과정 없이도 독립적으로 설계된 구조를 통해 그림자 영역의 공간적 정보를 그대로 보존하면서도 그림자 없는 영상을 생성해 내기 때문에, 기존의 신경망 기반 연구 대비, 계산상 장점을 가진다. 이는 표 2, 3을 통해 확인 가능하며, 표 4에서 정량적인 지표를 통해서도 이전의 연구 결과들 보다, 적어도 5dB 높은 PSNR값을 가지는 우수한 성능을 나타냄을 알 수 있다. 결론적으로 제안하는 알고리즘은 기존의 방식이 가진 제한점을 극복할 뿐만 아니라, 그림자와 물체의 경계면에 발생하는 고질적인 잡음 형태의 문제까지도 정교하게 제거할

수 있는 방법임이 확인되었다.

References

- [1] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," in *Proc. Comp. Vision, Graphics, and Image*, vol. 29, no. 1, pp. 100-132, Jan. 1985.
- [2] Z. Q. Zhao, et al., "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. and Learn. Syst.*, vol. 30, no. 11, pp. 3212-3232, Nov. 2019.
- [3] L. Xu, F. Qi, and R. Jiang, "Shadow removal from a single image," in *Proc. 6th Int. Conf. Intel. Sys. Design and Appl.*, vol. 2, pp. 1049-1054, Jian, China, Oct. 2006.
- [4] R. Guo, Q. Dai, and D. Hoiem, "Single-image shadow detection and removal using paired regions," in *Proc. IEEE Conf. Comp. Vision and Pattern Recog.*, pp. 2033-2040, Colorado Springs, USA, Jun. 2011.
- [5] J. Wei, et al., "Shadow inpainting and removal using generative adversarial networks with slice convolutions," *Comp. Graphics Forum*, vol. 38, no. 7, pp. 381-392, Nov. 2019.
- [6] H. Yoon, K. Kim, and J. Chun, "GAN-based shadow removal using context information," *J. Internet Comput. and Serv.*, vol. 20, no. 6, pp. 29-36, Dec. 2019.
- [7] V. Nguyen, et al., "Shadow detection with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Computer Vision*, pp. 4510-4518, Venice, Italy, Oct. 2017.
- [8] J. Wang, L. Xiang, and Y. Jian, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in *Proc. IEEE Conf. Comp. Vision, and Pattern Recog.*, pp. 1788-1797, Salt lake city, USA, Jun. 2018.
- [9] P. Isola, et al., "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comp. Vision and Pattern Recog.*, pp. 1125-1134, Hawaii, USA, Jul. 2017.
- [10] S. Ioffe and C. Szegedy, "Batch normalization:

Accelerating deep network training by reducing internal covariate shift,” in *Proc. Int. Conf. Mach. Learn.*, vol. 37, pp. 448-456, Male Bielice, Slovakia, May 2015.

- [11] L. Qu, et al., “Deshadownet: A multi-context embedding deep network for shadow removal,” in *Proc. IEEE Conf. Comp. Vision and Pattern Recog.*, pp. 4067-4075, Hawaii, USA, Jul. 2017.
- [12] X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in *Proc. IEEE Int. Conf. Computer Vision*, pp. 1501-1510, Venice, Italy, May 2017.
- [13] T. Park, et al., “Semantic image synthesis with spatially-adaptive normalization,” in *Proc. IEEE Conf. Comp. Vision, and Pattern Recog.*, pp. 2337-2346, Long Beach, USA, Jun. 2019.
- [14] H. Le and S. Dimitris, “Shadow removal via shadow image decomposition,” in *Proc. IEEE Int. Conf. Computer Vision*, pp. 8578-8587, Seoul, Korea, Oct. 2019.
- [15] Z. Pan, et al., “Recent progress on generative adversarial networks (GANs): A survey,” *IEEE Access*, vol. 7, pp. 36322-36333, Mar. 2019.
- [16] A. Jabbar, X. Li, and B. Omar, “A survey on generative adversarial networks: Variants, applications, and training,” *ACM Computing Surveys*, vol. 54, pp. 1-49, Oct. 2021.
- [17] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comp. and Computer-assisted Interv.*, pp. 234-241, Munich, Germany, Oct. 2015.
- [18] X. Cun, C. Pun, and C. Shi. “Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN,” in *Proc. AAAI Conf. Artif. Intel.*, vol. 34, pp. 10680-10687, New York, USA, Feb. 2020.

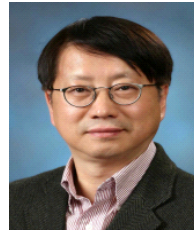
류 현 정 (Hyunjeong Ryu)



2010 8월 : 홍익대학교 전자·전기 공학부 졸업
 2020 3월~현재 : 연세대학교 전기전자공학과 석사과정
 <관심분야> 비디오/영상 신호 처리 및 해석, 딥러닝

[ORCID:0000-0002-9559-7756]

최 윤 식 (Yoonsik Choe)



1979년 2월 : 연세대학교 전기 공학과 졸업
 1984년 5월 : Case Western Reserve 대학, 시스템공학과 석사
 1987년 5월 : Pennsylvania 주립 대학 전기 및 컴퓨터공학과 석사

1990년 12월 : Purdue 대학 전기 및 컴퓨터공학과 박사
 1993년 3월~현재 : 연세대학교 전기전자공학과 교수
 <관심분야> 비디오/영상 신호 처리 및 해석, 통계적 영상신호처리

[ORCID:0000-0002-4856-8741]