

일반논문 (Regular Paper)

방송공학회논문지 제26권 제4호, 2021년 7월 (JBE Vol. 26, No. 4, July 2021)

<https://doi.org/10.5909/JBE.2021.26.4.419>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 자기 주의 증류를 이용한 심층 신경망 기반의 그림자 제거

김 진 희<sup>a)</sup>, 김 원 준<sup>a)†</sup>

### Shadow Removal based on the Deep Neural Network Using Self Attention Distillation

Jinhee Kim<sup>a)</sup> and Wonjun Kim<sup>a)†</sup>

#### 요 약

그림자 제거는 객체 추적 및 검출 등 영상처리 기술의 핵심 전처리 요소이다. 최근 심층 합성곱 신경망 (Deep Convolutional Neural Network) 기반의 영상 인식 기술이 발전함에 따라 심층 학습을 이용한 그림자 제거 연구들이 활발히 진행되고 있다. 본 논문에서는 자기 주의 증류(Self Attention Distillation)을 이용하여 심층 특징을 추출하는 새로운 그림자 제거 방법을 제안한다. 제안된 방법은 각 층에서 추출된 그림자 검출 결과를 하향식 증류를 통해 점진적으로 정제한다. 특히, 그림자 검출 결과에 대한 정답을 이용하지 않고 그림자 제거를 위한 문맥적 정보를 형성함으로써 효율적인 심층 신경망 학습을 수행한다. 그림자 제거를 위한 다양한 데이터셋에 대한 실험 결과를 통해 제안하는 방법이 실제 환경에서 발생한 그림자 제거에 효과적임을 보인다.

#### Abstract

Shadow removal plays a key role for the pre-processing of image processing techniques such as object tracking and detection. With the advances of image recognition based on deep convolution neural networks, researches for shadow removal have been actively conducted. In this paper, we propose a novel method for shadow removal, which utilizes self attention distillation to extract semantic features. The proposed method gradually refines results of shadow detection, which are extracted from each layer of the proposed network, via top-down distillation. Specifically, the training procedure can be efficiently performed by learning the contextual information for shadow removal without shadow masks. Experimental results on various datasets show the effectiveness of the proposed method for shadow removal under real world environments.

Keyword : shadow removal, shadow detection, convolutional neural networks, self attention distillation

a) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)

† Corresponding Author : 김원준(Wonjun Kim)

E-mail: wonjikim@konkuk.ac.kr

Tel: +82-2-450-3396

ORCID: <http://orcid.org/0000-0001-5121-5931>

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1F1A1068080).

※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1F1A1068080).

· Manuscript received April 28, 2021; Revised June 11, 2021; Accepted July 9, 2021.

Copyright © 2021 Korean Institute of Broadcast and Media Engineers. All rights reserved.

"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

## I. 서 론

최근 카메라 기술의 발전에 힘입어 다양한 모바일 기기를 이용하여 영상을 손쉽게 획득하게 되었다. 그러나, 촬영된 영상 내 다양한 조명 조건에 의해 그림자가 빈번하게 발생하며 이는 정확한 객체 검출에 많은 어려움을 준다 (그림 1 참조). 구체적으로, 그림 1의 (b) 영상에서 볼 수 있듯이 그림자 영역까지 함께 포함하여 객체 경계 상자(Bounding Box)가 검출되기 때문에 객체 추적 및 검출의 성능 하락을 초래하며, 그림 1의 (a)와 (c) 영상에서도 그림자가 영상의 품질 저하와 검출의 성능 저하를 초래하는 것을 알 수 있다. 따라서, 장면 인식 및 물체 추적 등 다양한 영상처리 기술의 원활한 구현을 위해 영상 내 그림자 영역을 검출하여 제거하는 연구는 매우 중요하다.

초기에는 그림자의 특성을 기반으로 사용자가 직접 설계한 특징(Handcrafted Feature)과 물리적 모델을 이용하여 그림자를 제거하는 연구가 진행되었다. 먼저 해당 방법들은 그림자 영역과 그림자가 제거된 영역에서 픽셀 간 선형 관계를 영상 형성 모델로 정의한다. 이후 물체의 색과 같은 다양한 특성을 활용해 영상 형성 모델의 매개변수를 추정하고, 그림자를 형성하는 매개변수의 값을 반전시켜 그림자 제거를 수행하였다<sup>[1][2]</sup>. 그러나 이러한 방법들은 영상 내 그림자와 유사한 색을 가지는 물체가 존재하거나 그림자의 배경이 복잡한 경우 그림자 제거의 성능이 저하되며, 전처리에 많은 수행 시간이 소요되는 문제점이 있다.

최근 그림자 검출 및 제거를 위한 데이터셋이 구축되고

있으며 합성곱 신경망(Convolution Neural Network, CNN) 기술이 발전하면서 영상 내 그림자를 제거하는 방법이 활발히 제안되고 있다. 심층 신경망 기반 방법은 그림자가 제거된 영상을 복원하기 위해 압축기-복원기(Encoder-Decoder) 구조를 이용하여 학습을 진행하며 그림자 영역에 속하는 픽셀이 영상 생성모델을 통해 효과적으로 제조명되도록 한다. 특히, 그림자 검출과 제거를 위해 두 개의 신경망을 동시에 학습하고 도출된 결과를 활용하는 ST-CGAN 방법이 최근 제안되었다<sup>[3]</sup>. ST-CGAN은 첫 번째 신경망으로부터 그림자 영역을 추출하고 이를 그림자 제거를 위한 신경망의 입력으로 함께 사용함으로써 복원 성능을 향상시켰다. 또한, 방향 별 가중치를 그림자 검출에 적용하여 그림자 제거 성능을 효과적으로 개선하는 방법이 제안되었다<sup>[4]</sup>. 방향 별 가중치 기반의 방법은 각 주요 방향(즉, 상, 하, 좌, 우)으로 인접 픽셀 간의 관계를 이용하여 공간적 맵 특징 지도를 학습하고, 시각 주의(Attention) 모듈을 각 방향별 특징 지도에 적용하여 효과적으로 그림자를 제거한다. 그러나 기존 방법들은 그림자 영역에만 집중하는 것이 아닌 영상 내 다양한 배경과 물체를 함께 고려하기 때문에 여전히 그림자 제거와 배경 색상 복원에 어려움을 겪고 있으며, 깊은 심층 신경망을 사용하여 많은 매개변수가 필요한 문제점이 있다.

이러한 문제점들을 해결하기 위해 본 논문에서는 자기 주의 종류를 이용한 심층 신경망 기반의 그림자 제거 방법을 제안한다. 제안하는 방법은 심층 신경망의 잠재 공간(Latent Space)에서 추출한 그림자 검출 결과와 입력 그림자 영상과의 차이를 통해 그림자 마스크를 획득하고, 이를

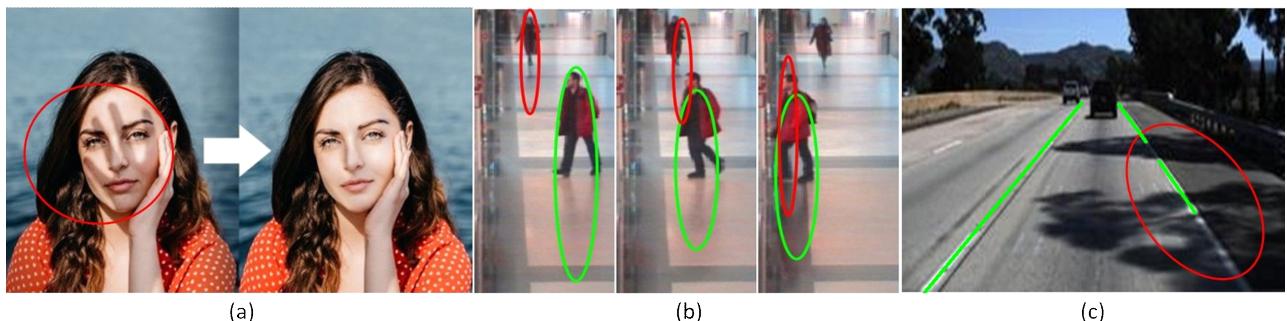


그림 1. 실생활 환경에서의 다양한 그림자 영상 예시. 그림자로 인한 다양한 컴퓨터 비전 기술의 성능 저하 예시. (a) 영상 품질 저하. (b) 객체 추적 성능 저하. (c) 차선 검출 성능 저하 (초록색 선은 예측된 차선을 나타냄)

Fig. 1. Examples of degradation performance for the techniques of computer vision by shadows. (a) degradation of image quality. (b) performance degradation of object tracking. (c) performance degradation of lane detection (green lines represent the estimated lanes)

다음 신경망 층에 삽입하여 마스크 영역에 대해 효과적인 그림자 제거를 수행한다. 특히, 신경망의 앞단에서 예측된 그림자 마스크는 하향식 증류(Distillation)를 통해 깊은 층에서 예측된 그림자 마스크가 더 정확하고 풍부한 문맥적 정보를 표현하도록 도와준다. 기존 방법과 달리 제안하는 방법은 그림자 정답 영상 없이 그림자 제거를 위한 문맥적 정보를 추출하여 적은 수의 가중치로 경쟁력 있는 성능을 달성하도록 하였다. 다양한 데이터셋에 대한 실험을 통해 제안하는 방법이 배경 색상에 대한 왜곡 없이 그림자 영역을 효과적으로 복원할 수 있음을 확인하였다. 본 논문의 구성은 다음과 같다. 2장에서는 제안하는 신경망 구조에 대해 자세히 설명하며, 3장에서는 다양한 실험을 통해 제안하는 방법이 기존 방법 대비 그림자 제거 성능이 뛰어남을 검증한다. 마지막으로 4장에서는 본 논문의 결론을 서술한다.

## II. 제안하는 방법

제안하는 방법은 신경망의 잠재 공간(Latent Space)에서 추출한 그림자 검출 결과를 계층별 하향식 증류를 통해 마스크 영역 정보를 깊은 계층에 전달함으로써 배경 색상에 대한 왜곡 없이 그림자를 효과적으로 제거한다. 또한, 제안하는 방법은 그림자 정답 영상을 이용한 별도의 학습을 수행하지 않기 때문에 기존 방법 대비 적은 수의 가중치로 그림자 제거를 수행한다. 본 장에서는 먼저 제안하는 신경망의 전체적인 구조를 소개한다. 이어서 그림자 마스크를 추출하기 위한 자기 주의 증류 과정에 대해 자세히 설명한 후 마지막으로 제안하는 구조를 기반으로 그림자 제거

학습에 사용된 손실 함수에 관해 설명한다.

### 1. 그림자 제거를 위한 제안하는 신경망의 전체적인 구조

제안하는 신경망의 전체적인 구조는 그림 2와 같다. ImageNet<sup>[5]</sup> 데이터셋을 기반으로 사전 훈련된 신경망 ResNext-101<sup>[6]</sup> 구조를 압축기(Encoder)로 사용하여 영상 특징을 다중 스케일에서 추출한다. VGG<sup>[7]</sup>와 잔류 밀도 네트워크(RDN)<sup>[8]</sup> 등과 같은 다른 합성곱 신경망도 제안된 신경망의 압축기로 사용될 수 있다. 압축기의 각 층(그림 2의 E1, E2, E3, 그리고 E4를 참조)에서 추출된 특징 지도는 세 개의 3×3 합성곱과 1×1 합성곱 계층(즉, 그림 2의 Conv Block)을 거쳐 입력 영상과 동일한 채널 개수를 가지는 그림자 제거에 관한 특징 지도를 생성한다. 제안하는 방법은 입력 그림자 영상을 다운 샘플링(Down Sampling)을 통해 그림자 제거를 위한 특징 지도와 동일한 해상도(Resolution)로 조정하고, 해상도가 조정된 입력 그림자 영상과 특징 지도의 픽셀별 차이를 구해 그림자 마스크를 생성한다. 모든 특징 스케일 공간의 압축기 단에서 위 과정을 반복하여 얻어진 그림자 마스크는 하향식 증류(즉, 그림 2의 빨간색 점선 화살표)를 통해 점진적으로 더 정확한 그림자 영역을 검출하도록 정제된다. 이후, 그림자 마스크와 다음 스케일의 압축기 단에서 추출한 특징 지도 간 픽셀 단위 합성곱 연산을 통해 그림자 영역이 강조된 특징 지도를 획득하여 다음 압축기 단의 입력으로 사용한다. 모든 특징 스케일 공간의 압축기 단에서 위 과정을 반복하여 얻어진 그림자 마스크는 하향식 증류를 통해 점진적으로 더 정확한 그

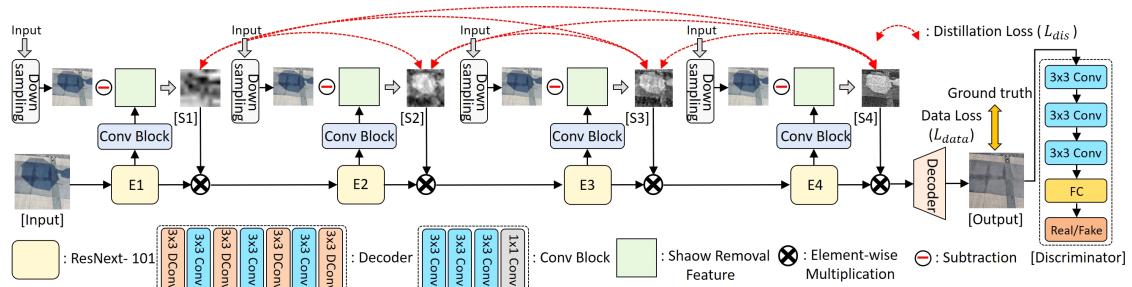


그림 2. 그림자 제거를 위한 제안하는 신경망의 전체적인 구조

Fig. 2. Overall architecture of the proposed network for shadow removal

그림자 영역을 검출하도록 정제된다. 압축기에서 최종적으로 추출된 특징 지도는 복원기(Decoder)를 통해 그림자가 제거된 영상으로 복원되며, 복원기의 구조는  $3 \times 3$  디콘볼루션(Deconvolution)과 세 개의  $3 \times 3$  합성곱으로 구성되어 있다. 복원된 그림자 제거 영상은 정답 영상(즉, 실제 그림자가 없는 영상)과 차이를 통해 손실 함수의 계산에 이용된다. 최종적으로 영상 전체의 자연스러운 색상 복원을 위해 판별기(Discriminator)를 이용한다. 판별기는 세 개의  $3 \times 3$  합성곱과 한 개의 완전 연결 층(Fully Connected Layer)으로 구성되어 있으며, 결과 값은 적대적 생성 신경망(Generative Adversarial Network, GAN)의 손실 함수 계산에 이용된다.

## 2. 자기 주의 종류(Self Attention Distillation)

초기의 지식 종류는 [12]에 의해 대규모의 심층 신경망에서 소규모의 심층 신경망으로 신경망의 학습 지식을 전달하기 위해 제안되었다. 일반적으로 지식 종류에서 교사 신경망과 학생 신경망은 동일한 모델의 용량을 공유하며, 동일한 차원의 신경망 층 간에 모방이 수행된다. 최근 방법들은 [13], [14] 지식 종류를 주의 종류의 개념으로 확장하여 해당 종류 과정을 심층 신경망 구조에 적용하였다. 특히, 제안하는 방법은 기존 주의 종류 방법들과 달리 교사 신경망이 따로 필요하지 않고, 하나의 신경망에서 생성된 주의 지도를 통해 주의 종류가 진행된다. 자기 주의 종류의 핵심은 신경망의 앞단에서 추출한 특정 정보를 깊은 층으로 전달하여 그림자의 풍부한 문맥적 정보를 추출하는 것이다. 그림 2에서 볼 수 있듯이 제안하는 자기 주의 종류 과정은 하향식 구조이며, E1, E2, E3, 그리고 E4 압축기 뒤에 검출

된 특징을 추가하여 각 스케일 별로 풍부한 그림자 특징을 활용할 수 있도록 한다. 또한, 기존의 하향식 자기 주의 종류의 흐름과 달리 조밀한(Dense) 연결로 주의(Attention) 특징 지도가 섬세하게 정제된다. 구체적으로, 첫 번째 압축기 층에서 자기 주의 종류 과정을 통해 추출된 그림자 마스크(S1)는 다음 층의 그림자 마스크(S2) 뿐만 아니라 세 번째와 네 번째 층의 그림자 마스크(S3와 S4)에도 문맥적 정보를 전달한다. 각 마스크 간 해상도(Resolution)가 다를 경우 이선형 업샘플링(Bilinear Upsampling)을 통해 동일한 해상도로 조정한다.

자기 주의 종류를 통해 추출한 그림자 주의 특징 지도의 시각화 결과는 그림 3에 나타내었다. 종류가 진행됨에 따라 그림자 영역이 효과적으로 강조되고 있음을 확인할 수 있다. 실험 결과 장에서 자기 주의 종류 과정을 적용한 경우와 적용하지 않은 경우에 대한 시각화 결과 비교 및 성능 비교를 수행하여 제안하는 자기 주의 종류에 대한 효과를 자세히 검증한다.

## 3. 그림자 제거를 위한 손실 함수 설계

본 논문에서는 제안하는 심층 신경망의 성능을 최적화하기 위해 기존 그림자 검출 및 제거에 널리 사용되는 손실 함수를 채택했다. 간단히 살펴보면, 신경망의 최종 예측 값에 적용되는 손실 함수  $L$ 은 예측된 그림자 제거 영상과 정답 영상(즉, 실제 그림자가 없는 영상) 간의 차이를 계산하기 위해 이용되는  $L_{data}$ , 그림자 마스크의 자기 주의 종류를 위해 이용되는  $L_{dis}$  및 적대적 생성 신경망(Generative Adversarial Network, GAN)의 손실 함수 항에 이용되는  $L_{adv}$ 의 합으로 정의되며 다음과 같이 표현할 수 있다.

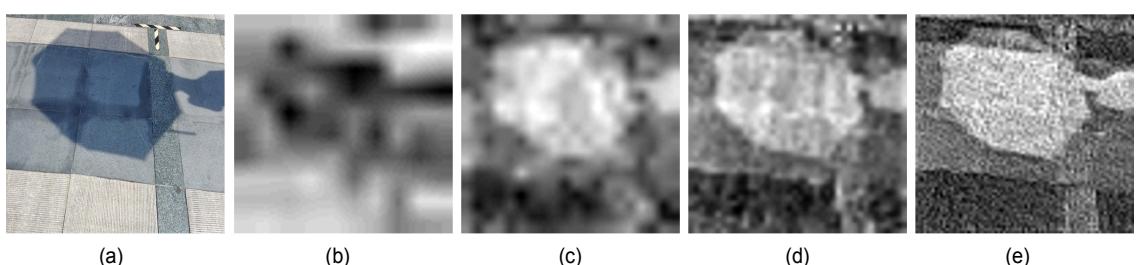


그림 3. 그림자 마스크(즉, 그림 2의 S1, S2, S3, 그리고 S4)의 시각화 결과. (a): 그림자 영상. (b)-(e): S1, S2, S3, 그리고 S4  
Fig. 3. Visualization results of shadow masks (i.e., S1, S2, S3, and S4). (a): shadow image. (b)-(e): S1, S2, S3, and S4

$$L = L_{data} + L_{dis} + L_{adv}, \quad (1)$$

여기서  $L_{data}$ 는 실제 그림자 제거 영상과 예측된 그림 제거 영상의 차이를 의미하며, 평균 제곱 오차(Mean Squared Error, MSE)를 이용하여 다음과 같이 계산된다.

$$L_{data} = \frac{1}{N} \sum_{i=1}^N (y_i - t_i)^2, \quad (2)$$

여기서  $y_i$ 는 예측된 그림자 제거 영상에서  $i$ 번째 픽셀의 값이며,  $t_i$ 는 실제 그림자 제거 영상에서  $i$ 번째 픽셀의 값이다.  $N$ 은 전체 픽셀의 개수를 나타낸다. 다음으로  $L_{dis}$ 는 각 신경망 층에서 추출된 그림자 마스크 간의 차이를 구하기 위해 이용되며, 아래와 같이 계산된다.

$$L_{dis} = \frac{1}{N} \left\{ \sum_{m=1}^{M-1} \sum_{k=m+1}^M (L_C(S^{(m)}, S^{(k)})) \right\}, \quad (3)$$

여기서  $S^{(m)}$ 과  $S^{(k)}$ 는 제안하는 신경망  $m$  번째와  $k$  번째 층에서 추출된 그림자 마스크를 의미한다.  $M$ 은 추출되는 그림자 마스크의 총 개수이며 제안한 방법에서는 4개가 사용되었다.  $L_C(\cdot, \cdot)$ 는 이진 교차 엔트로피(Binary Cross Entropy)를 나타내며 아래와 같이 계산된다.

$$L_C(p_i, q_i) = \sum_i -q_i \log(f(p_i)) - (1 - q_i) \log(1 - f(p_i)), \quad (4)$$

여기서  $q_i$ 는 그림자 정답 영상에서의  $i$ 번째 픽셀이 그림자 영역에 속할 확률을 나타내며,  $p_i$ 는 예측된 그림자 마스크 내  $i$ 번째 픽셀이 그림자 영역에 속할 확률 값이다.  $q_i$ 는 해당 픽셀이 그림자 영역인 경우 1의 값으로 할당되고, 그림자 영역이 아닌 경우 0으로 할당된다.  $f(\cdot)$ 는 시그모이드(Sigmoid) 함수를 의미한다. 마지막으로  $L_{adv}$ 는 영상 전체의 자연스러운 색상 복원을 위해 사용되며, 앞서 언급한 이진 교차 엔트로피를 기반으로 아래와 같이 계산된다.

$$L_{adv} = -\log(D(I_{real})) - \log(1 - D(G(I))), \quad (5)$$

여기서  $G(\cdot)$ 는 생성기로 제안하는 신경망을 나타내고,  $D(\cdot)$ 는 판별기를 나타낸다. 판별기의 결과 값은 정답 영상 여부에 대한 확률값을 나타낸다.  $I_{real}$ 은 실제 그림자가 없는 영상을 의미하며,  $I$ 는 신경망의 입력 영상으로 사용되는 그림자 영상을 나타낸다.

### III. 실험 결과 및 분석

본 논문에서 제안하는 방법의 성능 평가를 위해 두 개의 벤치마크 데이터셋을 사용하였다. 먼저, 그림자 제거 분야에서 일반적으로 사용되는 SRD 데이터셋<sup>[9]</sup>을 사용하였다. SRD 데이터셋은 다양한 조명과 배경에서 촬영된 그림자와 그림자 제거 영상의 묶음으로 구성되어 있으며, 최초의 그림자 제거를 위한 대규모 데이터셋이다. 학습을 위해 SRD 데이터셋은 총 2,680 장의 그림자와 그림자 제거 영상 묶음이 사용되었고, 408장의 묶음이 검증을 위해 사용되었다. 다음으로는 그림자 제거뿐만 아니라 그림자 검출에도 사용되는 ISTD 데이터셋<sup>[3]</sup>을 사용하였다. ISTD 데이터셋은 135개의 서로 다른 시나리오에서 촬영되었으며, 총 1,330 장의 그림자, 그림자 마스크, 그리고 그림자 제거 영상의 학습용 묶음과 540장이 검증용 영상 묶음으로 구성되어 있다. 일반적으로 그림자 제거를 위한 데이터셋은 먼저 그림자가 있는 장면의 영상을 촬영한 다음 그림자 생성을 유발하는 물체를 제거하여 그림자가 없는 영상을 촬영하는 방식으로 구축된다. 각 촬영 과정에서 광도와 카메라 노출이다를 수 있으므로, 그림자 제거 데이터셋의 입력 영상과 정답 영상의 쌍은 일관되지 않은 색과 광도를 가질 수 있다. 따라서, 이러한 일관되지 않은 훈련 영상 쌍으로 학습을 수행하게 되면 심층 신경망은 더 밝거나 더 어두운 편향된 결과를 생성할 수 있다. 이러한 문제점을 개선하기 위해 제안하는 방법과 기존 학습 기반 방법들은 입력 영상의 색상이 정답 영상의 색상과 동일하게 되도록 심층 신경망을 최적화하여 그림자 제거를 수행한다. 제안하는 방법은 PyTorch<sup>[10]</sup> 프레임워크에 기반하여 구현되었다. 본 논문에서 신경망 가중치를 최적화하기 위한 알고리즘으로 확률적 경사 하강법(Stochastic Gradient Descent, SGD)을 사용하

였고, 파워(Power)와 가속도(Momentum) 그리고 가중치 감쇠(Weight Decay) 값은 각각  $9 \times 10^{-1}$  와  $9 \times 10^{-3}$  그리고  $5 \times 10^{-4}$ 로 설정하였다. 학습 속도(Learning Rate)는  $5 \times 10^{-3}$ 부터 시작하여  $0.005 \times (1 - \text{현재 에포크(Epoch)}/\text{전체 에포크})$ 로 신경망을 학습하는 동안 감소하며, 전체 에포크는 500으로 설정하였다. 모든 영상은  $320 \times 320$  화소의 크기로 재조정되며, 과적합(Overfitting) 문제를 해결하기 위해 영상을 수직 및 수평으로 반전하는 데이터 증강 방법을 적용하였다. 제안하는 심층 신경망의 학습 소요 시간은 11시간이며, 학습과 성능 평가에는 Intel i7-68850K@3.60 GHz CPU와 NVIDIA GTX Titan XP GPU 2대가 이용되었다. 해당 조건에서의 그림자 제거 수행 속도는 평균 약 23.1fps이다.

제안하는 방법의 효율성을 검증하기 위해 본 논문에서는 두 개의 최신 그림자 제거 방법, 즉, Wang et al. [3]과 Hu et al. [4]와 성능을 비교하였다. 다양한 환경에서의 정성적

인 그림자 제거 결과를 그림 4와 5에 나타내었다. 특히, 제안하는 방법은 배경 색상의 큰 왜곡 없이 효과적으로 그림자를 제거함을 확인할 수 있으며, 복잡한 배경이 존재하는 영상에서도 효과적으로 그림자 제거를 수행한다 (그림 4와 5의 네 번째 줄 참조). 이러한 결과들을 바탕으로 제안하는 자기 주의 증류를 이용한 심층 신경망이 그림자 제거에 효과적임을 확인할 수 있다. 다음으로는 그림자 제거 성능의 정량적 평가에 널리 사용되는 평균 제곱근 편차(Mean Squared Error, RMSE)를 활용하여 기존 방법과 성능 비교를 수행하였으며, 해당 결과를 표 1과 2에 나타내었다. 평균 제곱근 편차는 LAB 색 공간에서 전체 영역, 그림자 영역, 그리고 그림자가 아닌 영역으로 나누어 계산된다. 표 1과 2에서 알 수 있듯이 제안하는 방법은 SRD와 ISTD 데이터셋에서 비교 방법 중 가장 높은 성능을 달성하였다. 그러나,

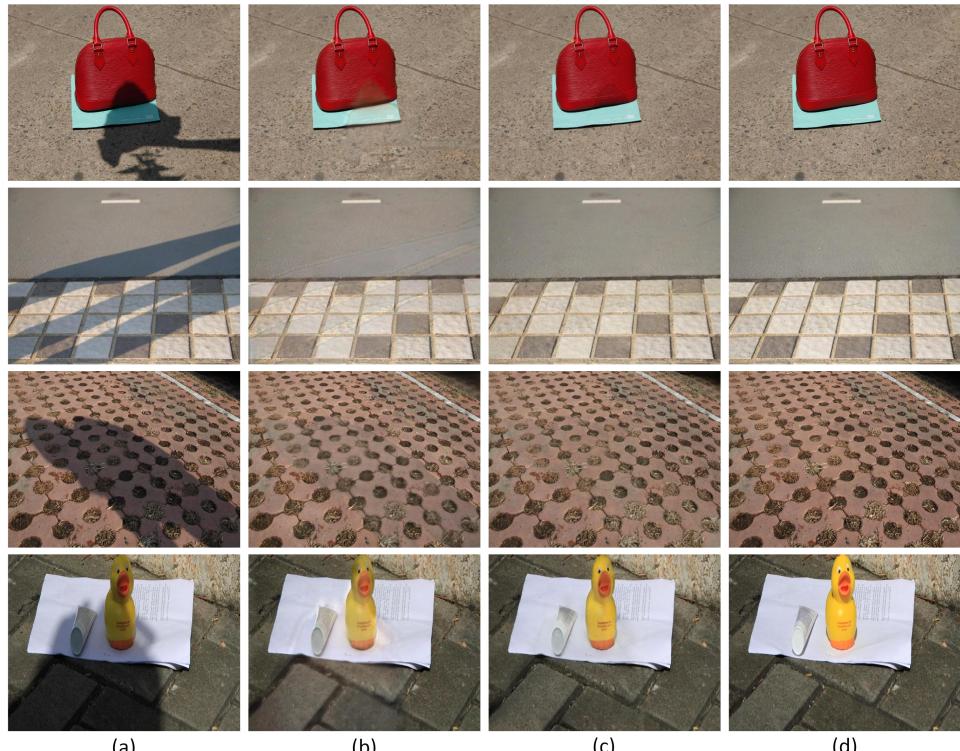


그림 4. SRD 데이터셋<sup>[9]</sup>에 대한 그림자 제거 결과. (a): 그림자 영상. (b)-(c): Hu et al. <sup>[4]</sup>과 제안하는 방법의 그림자 제거 결과. (d): 정답 영상

Fig. 4. Examples of shadow removal on SRD dataset<sup>[9]</sup>. (a): shadow images. (b)-(c): removal results by Hu et al. <sup>[4]</sup> and the proposed method. (d): ground truth

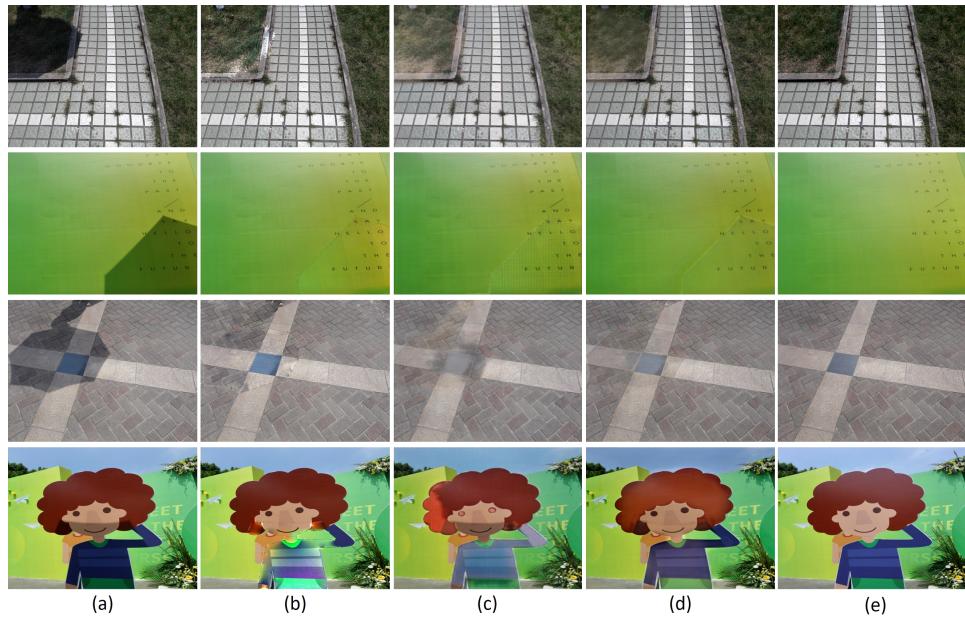


그림 5. ISTD 데이터셋<sup>[3]</sup>에 대한 그림자 제거 결과. (a) 그림자 영상. (b)-(d) Wang et al. [3], Hu et al. [4], 그리고 제안하는 방법의 그림자 제거 결과. (e) 정답 영상

Fig. 5. Examples of shadow removal on ISTD dataset<sup>[3]</sup>. (a) shadow images. (b)-(d) removal results by Wang et al. [3], Hu et al. [4], and the proposed method. (e) ground truth

표 1. SRD 데이터셋<sup>[9]</sup>에서의 정량적 평가 비교

Table 1. Quantitative evaluations on the SRD dataset[9]

Method	SRD dataset		
	Shadow	Non Shad.	All
Gong et al.[4]	19.58	4.92	8.73
DeshadowNet method	17.96	6.53	8.47
Hu et al.[4]	11.31	6.72	7.83
Proposed method	10.89	6.59	7.68

표 2. ISTD 데이터셋<sup>[3]</sup>에서의 정량적 평가 비교

Table 2. Quantitative evaluations on the ISTD dataset[3]

Method	ISTD dataset		
	Shadow	Non Shad.	All
VGG <sup>[7]</sup> - Wang et al. [3]	10.31	6.92	7.83
ResNext-101 <sup>[6]</sup> - Wang et al. [3]	9.84	6.71	7.48
VGG <sup>[7]</sup> - Hu et al. [4]	9.29	6.62	7.21
ResNext-101 <sup>[6]</sup> - Hu et al. [4]	9.22	6.50	7.10
VGG <sup>[7]</sup> - Proposed method	9.10	6.42	7.13
ResNext-101 <sup>[6]</sup> - Proposed method	8.89	6.27	6.95

제안하는 방법은 영상에서 그림자가 차지하는 비중이 작거나 그림자가 명확하게 표시되지 않으면 압축기의 하위 단으로 부족한 그림자 제거 정보가 전달되어 점진적으로 그림자를 제거하는 데 어려움이 있다 (예를 들어, 그림 7의 예시).

자기 주의 증류의 효과를 면밀히 검증하기 위해 자기 주의 증류 과정을 제외한 심층 신경망을 구성하여 비교 실험을 진행하였으며 그림 6과 표 3에 그 결과를 나타내었다. 해당 결과에서 볼 수 있듯이 자기 주의 증류 과정을 제외한 경우 그림자 제거 성능이 현저히 저하되며, 그림자 영역뿐만 아니라 배경 색상이 자연스럽게 복원되지 못하는 것을 확인할 수 있다. 따라서, 제안하는 자기 주의 증류 과정이 다양한 영상 내 그림자 영역을 제거하는 데 매우 효과적이라고 할 수 있다. 더 나아가, 제안하는 방법과 비교하는 방법 간의 매개변수 개수 비교를 진행하여 표 4에 나타내었다. 표 4에서 볼 수 있듯이 제안하는 방법이 기존 방법 대비 적은 수의 매개변수 개수로 우수한 성능을 나타내는 것을 확인할 수 있다. 마지막으로, 제안하는 지식 증류를 통한 이점을 상세히 알아보기 위해 지식 증류 손실 함수에 다양

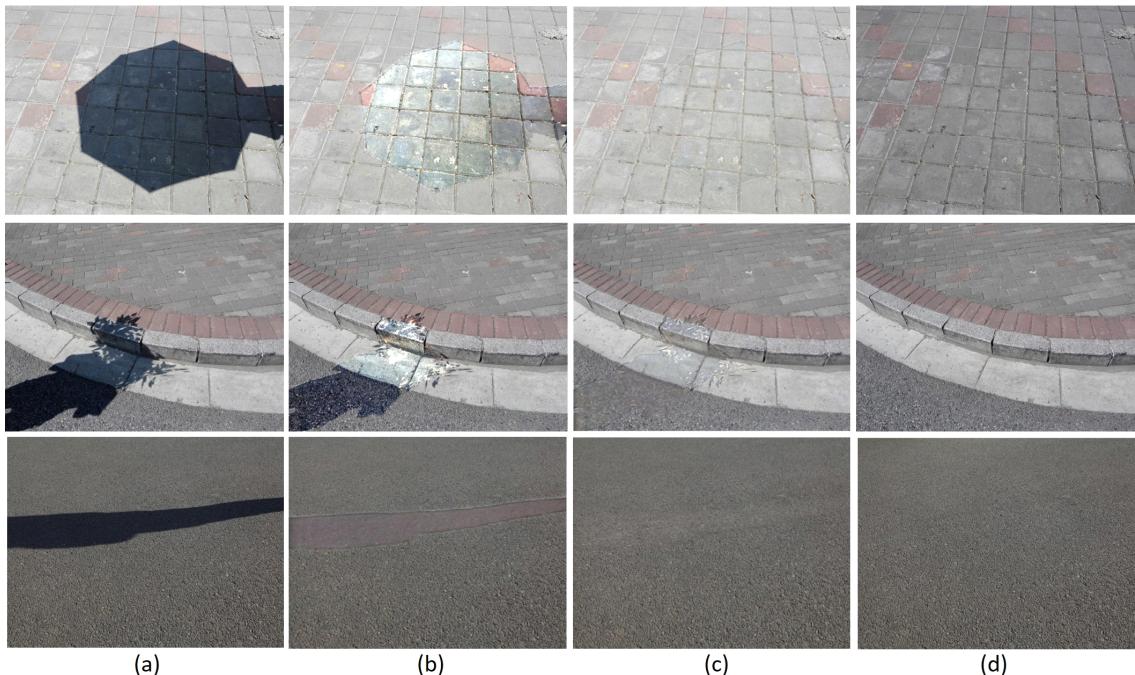


그림 6. 제안하는 심층 신경망 구조의 변화에 따른 그림자 제거 결과 비교. (a) 그림자 영상. (b) 자기 주의 종류가 없는 구조의 결과. (c) 자기 주의 종류가 있는 구조에서의 그림자 제거 결과. (d) 정답 영상

Fig. 6. Examples of shadow removal according to changes of the proposed network architecture. (a) shadow images. (b) removal results by model without self attention distillation (c) removal results by model with self attention distillation. (d) ground truth

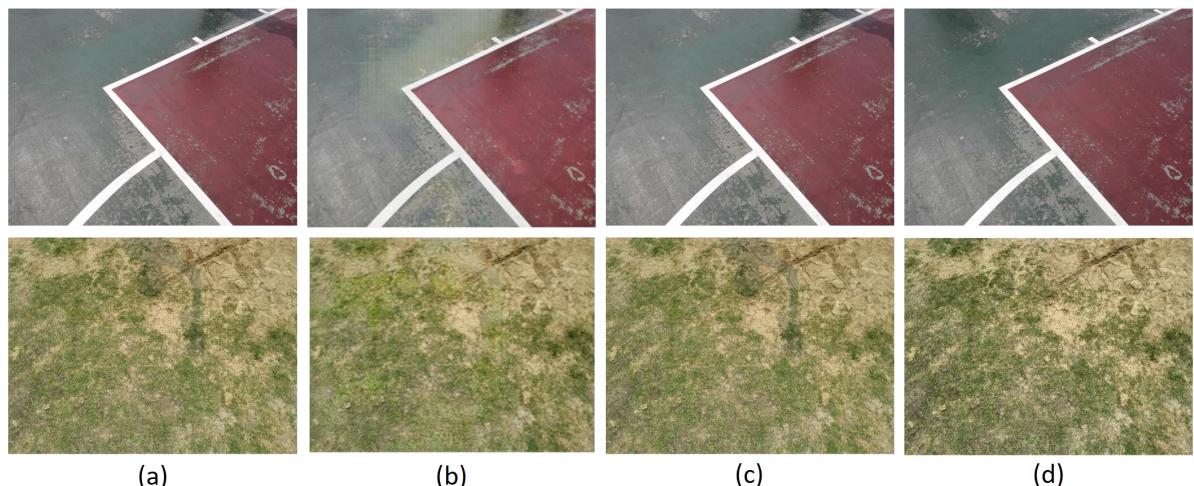


그림 7. 제안하는 방법의 그림자 제거 실패 예시. (a) 그림자 영상. (b)-(c) Hu et al. [4]과 제안하는 방법의 그림자 제거 결과. (d) 정답 영상  
Fig. 7. Examples of fail shadow removal of the proposed method. (a) shadow images. (b)-(c) removal results by Hu et al. [4] and the proposed method. (d) ground truth

한 하이퍼 파라미터 (Hyper Parameter) 값을 적용한 실험을 진행하여 표 5에 나타내었다. 표 5에서 알 수 있듯이 전체

손실 함수에서 종류 손실 함수의 비중이 높을수록 성능이 높아지는 것을 확인 할 수 있다.

표 3. 제안하는 심층 신경망 구조의 변화에 따른 정량적 실험 결과  
Table 3. Quantitative experimental results according to changes of the proposed network architectures

Method	SRD dataset		
	Shadow	Non Shad.	All
Full model (Ours)	10.89	6.59	7.68
w/o Self Attention Distillation	11.14	6.98	8.13
ISTD dataset			
Method	Shadow	Non Shad.	All
Full model (Ours)	8.89	6.72	7.83
w/o Self Attention Distillation	10.89	6.43	7.56

표 4. 제안하는 방법과 기존 방법 간의 매개변수 개수 비교  
Table 4. Comparison of the number of parameters between the proposed method and the existing methods

Model	Parameter (M)
VGG [7] - DeshadowNet et al. [9]	210.8
ResNext-101 [6] - DeshadowNet et al. [9]	131.9
VGG [7] - Wang et al. [3]	203.2
ResNext-101 [6] - Wang et al. [3]	122.8
VGG [7] - Hu et al. [4]	172.1
ResNext-101 [6] - Hu et al. [4]	103.4
VGG [7] - Proposed method	155.3
ResNext-101 [6] - Proposed method	88.6

표 5. 제안하는 증류 손실 함수의 하이퍼 파라미터 (Hyper Parameter) 값 적용에 따른 그림자 제거 성능 비교  
Table 5. Comparison of the performance for shadow removal according to hyperparameter values of the proposed distillation loss function

Dataset	Hyper Parameter (M)			
	0.05	0.1	0.5	1 (Ours)
ISTD	7.49	7.24	7.03	6.95

#### IV. 결 론

본 논문에서는 자기 주의 증류를 이용한 심층 신경망 기반의 그림자 제거 방법을 제안하였다. 제안하는 방법은 자기 주의 증류 과정을 통해 심층 신경망의 잠재 공간(Latent Space)에서 추출한 그림자 마스크를 정제하고, 해당 마스크

를 이용해 배경 색상에 대한 큰 왜곡 없이 그림자 영역을 복원한다. 특히, 다중 스케일 내 압축기 단에서 추출한 풍부한 영상 특징들을 조밀한 하향식 증류(Distillation)를 통해 깊은 층에 전달함으로써 예측된 그림자 마스크가 풍부한 문맥적 정보를 표현하도록 한다. 또한, 제안하는 방법은 그림자 정답 영상 없이 그림자 제거를 위한 학습을 수행하여 적은 수의 매개변수로 경쟁력 있는 성능을 달성하였다. 다양한 실험을 통해 제안하는 방법이 기존 방법 대비 그림자 제거에 효과적임을 확인하였다.

#### 참 고 문 헌 (References)

- [1] G. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59 - 68, Nov. 2006.
- [2] R. Guo, Q. Dai, and D. Hoiem, "Paired regions for shadow detection and removal," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2956 - 2967, Oct. 2013.
- [3] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2018, pp. 1788 - 1797.
- [4] X. Hu, C. Fu, L. Zhu, J. Qin, and P. Heng, "Direction-aware spatial context features for shadow detection and removal," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 11, pp. 2795-2808, May. 2019.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Advances in Neural Information Processing Systems, Dec. 2017, pp. 1097 - 1105.
- [6] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2017, pp. 1493 - 1500.
- [7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Representation, May 15, 2015, pp. 1 - 14.
- [8] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2018, pp. 2472 - 2481.
- [9] L. Qu, J. Tian, S. He, Y. Tang, and R. W. Lau, "DeshadowNet: A multi-context embedding deep network for shadow removal," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2017, pp. 4067 - 4075.
- [10] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. Devito, Lin, A. Desmaison, L. Antiga, A. Lerer, "Automatic differentiation in pytorch", in Proc. Conference and Workshop on Neural Information Processing Systems (NIPS), pp. 1-4, 2017.

- [11] H. Gong and D. P. Cosker, "Interactive shadow removal and ground truth for variable scene categories," in Proc. British Machine Vision Conference, Sep. 2014, pp. 1 - 11.
- [12] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in Proc. Conference and Workshop on Neural Information Processing Systems (NIPS), Dec. 2015, pp. 1-9.
- [13] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, "Residual attention network for image classification," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2017, pp. 3156 - 3164.
- [14] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer," in Proc. International Conference on Learning Representations, Sep. 2017, pp. 1 - 13.

---

### 저자 소개

---

#### 김 진 희



- 2020년 2월 : 건국대학교 학사
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 석사과정
- ORCID : <https://orcid.org/0000-0001-6061-5363>
- 주관심분야 : 컴퓨터 비전, 객체 검출, 기계학습, 패턴 인식

#### 김 원 준



- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성중합기술원 전문연구원
- 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
- ORCID : <https://orcid.org/0000-0001-5121-5931>
- 주관심분야 : 영상이해, 컴퓨터 비전, 기계학습, 패턴 인식