

일반논문 (Regular Paper)

방송공학회논문지 제25권 제6호, 2020년 11월 (JBE Vol. 25, No. 6, November 2020)

<https://doi.org/10.5909/JBE.2020.25.6.980>

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

다양성 및 안정성 확보를 위한 스타일 전이 네트워크 손실 함수 정규화 기법

최인성^{a)}, 김용구^{a)†}

A Normalized Loss Function of Style Transfer Network for More Diverse and More Stable Transfer Results

Insung Choi^{a)} and Yong-Goo Kim^{a)†}

요 약

딥-러닝 기반 스타일 전이 기법은 영상의 고차원적 구조적 특성을 적절하게 반영하여 높은 품질의 스타일 전이 결과를 제공함으로써 최근 크게 주목받고 있다. 본 논문은 이러한 딥-러닝 기반 스타일 전이 방식의 안정적이고 보다 다양한 스타일 전이 결과 제공에 대한 문제를 다룬다. 스타일 전이를 위한 광범위한 초-매개변수 설정에 따른 실험 결과에 대한 고찰을 바탕으로 스타일 전이 결과의 안정성 및 다양성에 대한 문제를 정의하고, 이러한 문제를 해결하기 위한 부분 손실 정규화 방법을 제안한다. 제안된 정규화 방식을 이용한 스타일 전이는 입력 영상의 특징에 상관없이 초-매개변수 설정을 통해 동일 수준의 스타일 전이 정도를 조절할 수 있을 뿐 아니라, 스타일 손실을 정의하는 계층 별 가중치 설정의 조절을 통해 기존 방식과 달리 보다 다양한 스타일 전이 결과를 제공하며, 입력 영상의 해상도 차이에 대해 보다 안정적인 스타일 전이 결과를 제공하는 특징을 가진다.

Abstract

Deep-learning based style transfer has recently attracted great attention, because it provides high quality transfer results by appropriately reflecting the high level structural characteristics of images. This paper deals with the problem of providing more stable and more diverse style transfer results of such deep-learning based style transfer method. Based on the investigation of the experimental results from the wide range of hyper-parameter settings, this paper defines the problem of the stability and the diversity of the style transfer, and proposes a partial loss normalization method to solve the problem. The style transfer using the proposed normalization method not only gives the stability on the control of the degree of style reflection, regardless of the input image characteristics, but also presents the diversity of style transfer results, unlike the existing method, at controlling the weight of the partial style loss, and provides the stability on the difference in resolution of the input image.

Keyword: Deep Learning, Style Transfer, VGG-19, Hyper-Parameters, Regularization

1. 서론

스타일 전이(Style Transfer)는 하나의 영상이 가지고 있는 질감, 색상, 스트로크 등 해당 영상을 인식하는데 필수적인 특징('스타일')을 다른 영상의 한 부분 혹은 영상 전체에 적용하는 방법을 의미하며, 텍스처 합성 혹은 텍스처 전이 방법의 일환으로 연구되었다^[1-2]. 전통적인 초기 연구들은 선형 필터 출력에 대한 히스토그램 매칭^[3] 또는 비-매개변수(Non-Parametric) 샘플링 기법^[4] 등, 주로 낮은 수준의 통계적 특징을 사용하여 영상의 텍스처를 모델링함으로써 스타일 전이를 수행했지만, 이로 인해 영상이 가진 높은 수준의 구조적 특징을 전이하는데 종종 그 한계를 보여 왔다. 최근 소개된 Gatys 등의 연구^[5]는, 합성곱 신경망(Convolutional Neural Network)의 출력이 가진 통계적 특징을 매칭하는 방식으로, 이전 기법들과 대별되는 새로운 스타일 전이 방법을 제시하였는데, 영상의 고차원적 스타일 구조를 적절히 반영한 매우 높은 품질의 스타일 전이 결과를 제공함으로써 크게 주목받고 있다. Gatys 등의 이러한 연구 결과를 기반으로 많은 후속 연구들이 추진되고 있는데, 전이 영상의 스트로크 크기를 조절하는 연구^[6], 스타일 전이의 대상 영역 지정 및 스케일 등을 조절하는 연구^[7], 스타일 전이의 수렴 현상 특징에 대한 연구^[8], 비디오 및 가상현실 미디어에 대한 스타일 전이 연구^[9-10] 등이 대표적이다. 특히, [5]의 느린 스타일 전이 속도 문제를 해결하기 위한 시도들이 매우 다양하게 시도되고 있는데, 전방 전달 신경망(Feedforward Neural Network)을 이용해 특정 스타일의 전이 손실 함수를 최적화 하는 연구^[11-12], 다중 스타일에 대한 전이를 하나의 신경망으로 처리하는 연구^[13], 그리고 다양한 스타일 입력에

대한 빠른 스타일 변환을 수행하는 연구^[14-16] 등이 있다. 이러한 다양한 인공 신경망 기반 스타일 전이 연구 중에서, [5]의 방식이 가장 다양하고 우수한 전이 품질을 제공할 수 있는 기법 중 하나로 알려져 있어, 본 논문에서는 [5]의 안정적 스타일 전이 문제에 관심을 집중하기로 한다.

Gatys 등의 스타일 전이 기법^[5]은 영상 인식에 사용되던 [17]의 합성곱 신경망을 이용한 최적화 문제로 정의되었는데, 반영할 스타일을 가진 영상의 신경망 출력 특징 벡터를 이용한 채널 간 상관도와, 스타일을 반영하고자 하는 대상 영상의 신경망 출력 특징 벡터의 유사도를 사용하여 손실 함수를 정의하고, 출력 영상에 대한 손실 역전파를 이용하여 정의된 손실 함수를 최적화하도록 출력 영상을 조절함으로써 스타일 전이가 수행되도록 하였다. 하지만, 이러한 손실함수의 정의는 사용된 신경망의 출력 특징에 따른 최적화 영향을 적절하게 고려하지 못함으로써, 스타일 전이의 특징 및 그 정도를 결정하는 초-매개변수(hyper-parameter)의 설정이 직관적이지 못할 뿐 아니라, 경우에 따라 전이 특성을 거의 조절할 수 없게 되거나 또는 입력 영상에 따라 달라질 수 있는 파라미터 범위를 벗어나는 경우 적절하지 못한 전이 결과를 얻게 되는 문제가 있다 (그림 2 참조). 이와 같은 문제를 해결하기 위해, 본 논문에서는, [5]에서 정의한 손실 함수의 편미분 과정을 분석하고, 이를 통해 신경망 특징에 따른 적절한 정규화 방법을 고안함으로써 다양한 입력 영상에 대해 안정적으로 스타일 전이를 수행할 수 있는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 본 논문의 2장에서는, [5]의 스타일 전이 방식과 그 최적화 과정을 상세히 설명하고, 이를 통해 초-매개 변수의 선택에 따른 스타일 전이의 불안정 요소 및 제한된 다양성 문제를 살펴본다. 이 후, 본 논문의 3장에서는 이러한 문제가 발생하는 원인과 이를 해결하기 위한 개선된 제안 손실 함수 설계를 상술하고, 4장에서 다양한 실험 영상에 대한 스타일 전이 안정성 및 다양성 문제에 대한 모의실험 결과를 보인다. 마지막으로, 5장에서 본 논문의 결론을 맺도록 한다.

a) 서울미디어대학원대학교 인공지능응용소프트웨어학과(Dept. of AI Software Eng., Seoul Media Institute of Technology)

‡ Corresponding Author : 김용구(Yong-Goo, Kim)
E-mail: ygkim@smit.ac.kr
Tel: +82-2-6393-3222

ORCID: <https://orcid.org/0000-0002-8905-1984>

※ This research is supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R202004 0238)

· Manuscript received October 6, 2020; Revised November 16, 2020; Accepted November 16, 2020.

II. 최적화 기반 스타일 전이의 안정성 및 다양성 문제

참고문헌 [5]의 스타일 전이 방식은 영상 인식을 위해 개발된 [17]의 합성곱(Convolution) 신경망(Neural Network)을 사용하고 있는데, 그림 1에 이러한 VGG-19(Visual Geometry Group-19) 신경망의 구조 및 특징을 도시하였다.

그림의 VGG-19 신경망은 [18]에 소개된 합성곱 신경망의 전형적인 구조에 따라, 5개의 합성곱 연산 그룹과 2개의 전-결합(FC - Fully Connected) 계층에 이은 출력 계층으로 구성되어 있다. 각 합성곱 그룹에는 다수의 합성곱 계층이 포함되는데, 그림에 이를 ConvX_Y (X: 그룹 번호, Y: 계층 번호)로 표시하였고, 각 합성곱 계층의 출력에는 정류된 선형 유닛(ReLU - Rectified Linear Unit) 활성화 함수가 적용된다 (ReLU_X_Y로 표시하며, 그림에서는 생략됨). 그림에 표현된 각 합성곱 계층 표시 옆의 괄호 안 두 숫자 K-C(예를 들어, Conv1_1(3-64)에서 3-64)는 합성곱 커널의 크기가 KxK이고 커널의 개수가 C개로 이루어진 계층임을 나타낸다.

[5]의 스타일 전이는 이러한 VGG-19 신경망에서 합성곱 계층 Conv5_4까지를 사용하여 수행되는데, 그 세부 과정을 설명하기 위해 s 를 스타일 영상(전이할 스타일을 포함하고 있는 $H_s \times W_s \times 3$ 크기의 RGB 영상), o 를 오브제 영상(스타일을 반영할 $H_o \times W_o \times 3$ 크기의 RGB 영상), 그리고 p 를 출력 영상(o 에 s 의 스타일을 반영한 $H_o \times W_o \times 3$

크기의 RGB 영상)이라 하면, p 는 다음을 만족하는 영상으로 정의된다.

$$p = \min_x L(o, s, x), \tag{1}$$

여기서, x 는 $H_o \times W_o \times 3$ 크기를 가지는 임의의 영상을 의미하며, $L(\cdot)$ 은 세 영상 o, s, x 로부터 정의되는 손실 함수로 식 (2)와 같이 오브제 손실 함수, $L_o(o, x)$,와 스타일 손실 함수, $L_s(s, x)$, 의 가중 합으로 정의된다.

$$L(o, s, x) = \lambda_{os} \cdot L_o(o, x) + L_s(s, x), \tag{2}$$

여기서 가중치 λ_{os} 는 스타일 손실에 대한 오브제 손실의 중요도를 의미하는 초-매개변수로 스타일 전이의 정도를 조절하는 파라미터이고, 영상 x 를 그림 1의 신경망 ϕ 에 입력하여 얻게 되는 계층 l 의 출력을 $\phi_l(x)$ 로 나타내면, 오브제 손실 함수와 스타일 손실 함수는 다음과 같이 정의된다.

$$L_o(o, x) = \sum_{j \in C} \frac{1}{2} \cdot \lambda_o(j) \cdot \|\phi_j(o) - \phi_j(x)\|_2^2, \tag{3}$$

$$L_s(s, x) = \sum_{j \in S} \left(\frac{1}{2H_j W_j C_j} \right)^2 \cdot \lambda_s(j) \cdot \|G(\phi_j(s)) - G(\phi_j(x))\|_F^2 \tag{4}$$

여기서, C 와 S 는 각기 오브제 손실과 스타일 손실의 정

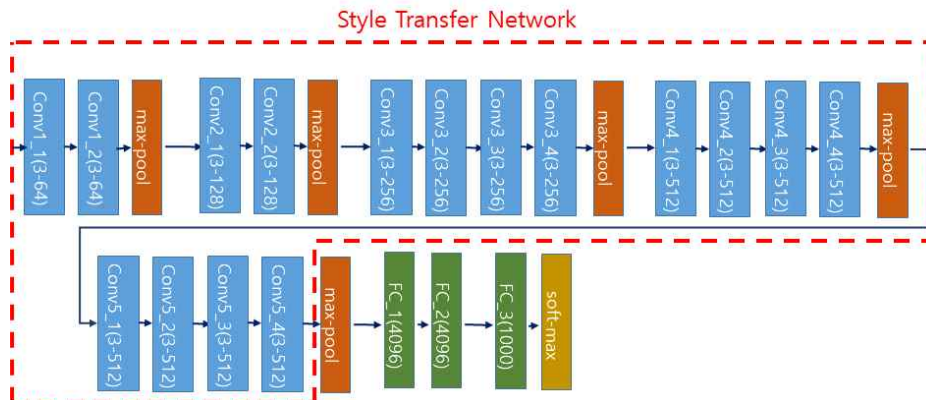


그림 1. VGG-19 네트워크의 구조 및 스타일 변환 네트워크 [17]
 Fig. 1. The structure of VGG-19 and style transfer network [17]

의에 사용되는 신경망 ϕ 의 계층 집합¹⁾을 의미하고, $\|\cdot\|_2$ 와 $\|\cdot\|_F$ 는 각기 2-노름(Norm), 그리고 프로베니우스(Frobenius) 노름을 의미한다. 또한, 식(4)에서 H_j, W_j, C_j 는 텐서 $\phi_j(x)$ 의 차원을 나타내는 파라미터로 $j = ConvX_Y$ (또는 $j = ReLuX_Y$)인 경우, $H_j = H_x/2^{X-1}$, $W_j = W_x/2^{X-1}$, $C_j = 64 \times 2^{X-1}$ 과 같고, $G(\phi_j(\cdot))$ 는 신경망 출력 $\phi_j(\cdot)$ 에 대한 $C_j \times C_j$ 크기의 그램-행렬(Gram Matrix)을 의미하는 것으로 다음과 같이 정의된다.

$$G(\phi_j(y)) = \Phi_j^T \Phi_j, \quad (5)$$

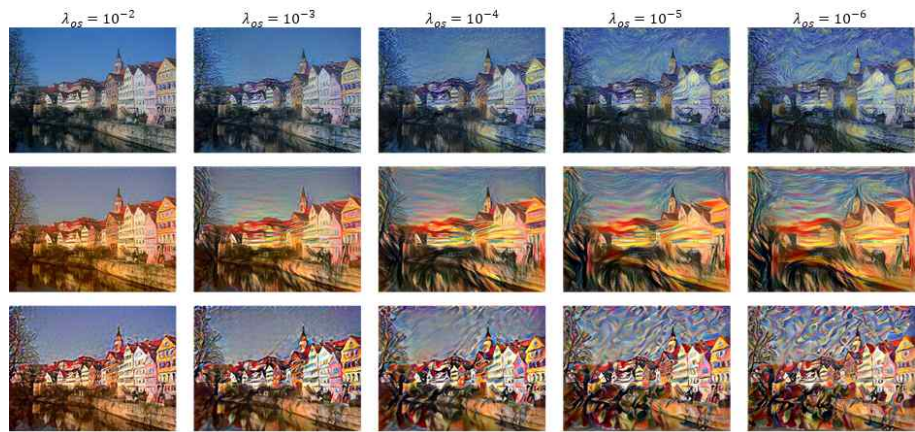
여기서, Φ_j 는 $H_j \times W_j \times C_j$ 크기의 텐서 $\phi_j(y)$ 를 $H_j W_j \times C_j$ 크기로 재구성한 행렬을 의미하며, $\phi_j(y)$ 의 각 C_j 차원 특징 벡터들을 행-벡터로 정렬한 행렬에 해당한다. 따라서 $G(\phi_j(y))$ 는 이러한 행-벡터들에 대한 $C_j \times C_j$ 크기의 그램-행렬이 된다. 마지막으로, 식 (3)의 $\lambda_o(j)$ 와 식 (4)의 $\lambda_s(j)$ 는 각기 오브제 손실과 스타일 손실의 정의에서 각 계층의 부분 손실이 가진 중요도를 설정하는 초-매개변수로, [19-20] 등의 연구 결과를 참고하면, 하위 계층의 가중치를 높여 스타일 영상의 국부적 특징을 우세하게 반영하거나, 상위 계층의 가중치를 높여 스타일 영상의 구조적 특징을 강하게 반영하는 역할을 한다.

이와 같이 설계된 [5]의 스타일 전이에 대한 기본적인 최적화 실험 결과를 그림 2에 도시하였다 (자세한 내용은 4장 실험 및 결과 참조). 그림 2.(a)는 서로 다른 3가지 스타일 영상을 사용하여 가중치 λ_{os} 의 변화에 따른 스타일 전이 결과를 도시한 것이다. 그림에서 확인할 수 있는 것은 λ_{os} 가 작아지면서 점차 스타일이 강하게 반영된다는 사실(가중치가 가진 본래의 의미에 부합)과, 어느 이상 스타일 전이를

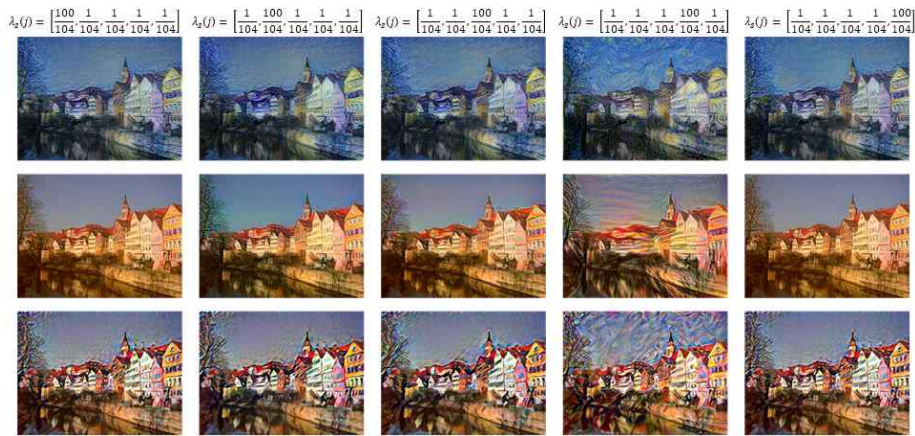
강하게 할 수 없다는 사실²⁾, 그리고 스타일 영상에 따라 적절한 스타일의 반영³⁾을 위한 가중치 값이 크게 차이가 난다는 점이다.

두 번째로, 그림 2.(b)는 $S = \{Conv1-2, Conv2-2, Conv3-4, Conv4-4, Conv5-4\}$ (식 (4) 참고)에 대한 $\lambda_s(j)$ 가중치 조절을 통해 수행된 스타일 전이 최적화 결과를 도시한 것이다. 그림 2.(b)에서 j번째 열에 위치시킨 그림들은 j번째 스타일 부분 손실이 다른 계층의 부분 손실에 비해 100배 만큼 크게 설정한 경우의 최적화 결과이다⁴⁾. 그림에서 확인할 수 있는 것은 4번째 계층(Conv4_4)에 대한 부분 손실을 강화한 경우를 제외하고는 모두 유사한 품질의 결과를 보이고 있으며, 특히 그 결과가 적절한 스타일 반영으로 볼 수 없다는 점이다. 100배 이외 다른 비율로 특정 계층의 부분 손실을 강화하는 경우, 1배(동일한 비율 - 그림 2.(a)와 같은 결과)에서 그 배수를 점차 증가시키면 4번째 계층을 강화한 결과를 제외하고는 점차 스타일 반영이 약화되다가 100배 이상이 되면 그림 2.(b)와 동일한 결과를 도출하게 된다. 4번째 계층을 사용한 부분 손실을 점차 강화한 결과에서만 그림 2.(b)에서 볼 수 있는 바와 같이 약간의 스타일 강화가 나타나는 것을 볼 수 있다. 마지막으로 그림 2.(c)는 적절한 스타일 반영이 가능한 가중치 조합을 사용하여, 각기 다른 해상도로 변환한 동일 영상에 대해 스타일 전이 실험을 수행한 결과를 도시한 것이다. 그림에서 확인할 수 있는 사실은, 특정 해상도(그림에서 512)에서 적절한 스타일 반영이 가능한 초-매개변수 조합이 낮은 해상도의 영상에 대한 스타일 전이에서는 과도한 스타일 전이 결과를 보이고, 높은 해상도의 영상에서는 매우 약한 스타일 전이 결과를 제공한다는 점이다. 이상에서 살펴본 바와 같이, [5]의 스타일 전이 알고리즘은 동일 파라미터를 사용하는 경우에도 해상도에 따라 서로 다른 스타일 반영 특성을 보이고,

1) 각 계층 집합은 다양하게 설정할 수 있으나, 쉬운 이해를 위해 다음과 같은 전형적인 예를 염두에 둘 수 있다.
 $C = \{ReLu4_2\}$, $S = \{ReLu1_1, ReLu2_1, ReLu3_1, ReLu4_1, ReLu5_1\}$ $S = \{ReLu1_1, ReLu2_1, ReLu3_1, ReLu4_1, ReLu5_1\}$
 2) 그림에 도시된 가중치 영역 밖에서는 모두 그 이전 결과와 동일한 결과가 얻어진다. 즉, 가중치를 더욱 낮게 설정하여 0으로 두는 경우에도, 그림 2.(a)의 오른쪽 끝 열에 주어진 그림과 동일한 결과를 얻게 된다.
 3) ‘적절한 스타일 반영’에 대한 느낌은 매우 주관적인 요소이나, 본 논문에서는 오브제 영상이 가진 객체 특성을 확인할 수 있는 수준에서 스타일 영상의 특성이 최대한 반영된 정도를 그 정의로 사용한다. 저자의 주관적 판단에 따른 그림 2.(a) 결과의 적절한 스타일 반영 파라미터는, 각 열에 대해 $10^{-5} \leq \lambda_{os} \leq 10^{-6}$, $10^{-3} \leq \lambda_{os} \leq 10^{-4}$, 그리고 $10^{-4} \leq \lambda_{os} \leq 10^{-5}$ 정도에 해당하는 것으로 보인다.
 4) 그림 2.(a)의 실험에서 $\lambda_s(j)$ 는 모두 동일한 값 0.2로 설정하였으며, 그림 2.(b)의 실험에서 각 열은 적절한 스타일 반영이 가능한 가중치 $\lambda_{os}(10^{-5}, 10^{-3}, 10^{-4})$ 를 사용한 결과이다.



(a) Test Results with λ_{os}



(b) Test Results with $\lambda_s(j)$



(c) Test Results with Image Size

그림 2. 초-매개변수 변화에 따른 기존 스타일 전이 기법[5]의 실험 결과
 Fig. 2. Simulation results of the style transfer [5] with the changes of hyper-parameters

가중치 선택에 따른 스타일 전이 영향이 특정 범위에서 매우 제한적으로 나타나며, 가중치의 선택에 대해 직관적으로 유추할 수 있는 스타일 전이 결과와 매우 상이한 결과가 나타나는 문제가 있다. 따라서 서로 다른 입력 영상으로부터 보다 안정적으로 다양한 스타일 전이 결과를 얻을 수 있는 방법에 대한 연구는 매우 자연스런 스타일 전이 네트워크 능력 확장의 문제라 할 수 있으며, 본 논문에서는 이러한 문제를 다루기로 한다.

III. 스타일 전이 네트워크 손실 함수 조절 기법

본 장에서는, 앞서 설명한 스타일 전이 네트워크의 정성적 실험 결과에 대한 이유를 분석하고, 이를 통해 초-매개변수 $\lambda_s(j)$ 의 선택에 따른 다양한 스타일 전이 결과를 서로 다른 입력 영상에서 유사한 범위의 초-매개변수 λ_{os} 의 선택을 통해 안정적으로 생성할 수 있는 손실 함수 정규화 기법을 제안한다.

1. 스타일 전이 네트워크의 수렴 특성

식 (1)로 주어지는 신경망 기반 스타일 전이 문제는 출력 영상에 대한 손실 함수 변화율을 기반으로 출력 영상을 반복 갱신하는 최적화 방법을 사용하여 해결할 수 있다. 즉, 현재의 출력 영상을 $p^{(t)} = \{p_{i,j,k}^{(t)} | 1 \leq i \leq H_o, 1 \leq j \leq W_o, 1 \leq k \leq C_o\}$ 라 하면, 경사 하강법(Gradient Descent)에 의해 갱신되는 다음 번 출력 영상은 다음과 같다.

$$p^{(t+1)} = p^{(t)} - \alpha^{(t)} \cdot g^{(t)}, \quad (6)$$

$$g^{(t)} = \frac{\partial L(s, c, p^{(t)})}{\partial p^{(t)}} = \left\{ \frac{\partial L(s, c, p^{(t)})}{\partial p_{i,j,k}^{(t)}} \mid i = 1, \dots, H_o, j = 1, \dots, W_o, k = 1, \dots, C_o \right\} \quad (7)$$

$$\frac{\partial L_s(s, p^{(t)})}{\partial p_{i,j,k}^{(t)}} = \sum_{l \in S} \frac{\lambda_s(l)}{2} \left(\frac{1}{H_l W_l C_l} \right)^2 \sum_{m,n} [G_{m,n}(\phi_l(s)) - G_{m,n}(\phi_l(p^{(t)}))] \cdot \left(\frac{\partial G_{m,n}(p^{(t)})}{\partial p_{i,j,k}^{(t)}} \right) \quad (10)$$

여기서 $\alpha^{(t)}$ 는 현재 출력 영상의 갱신을 위한 학습율을 의미하고, $g^{(t)}$ 는 현재 출력 영상 $p^{(t)}$ 에 대한 식 (2)의 손실 함수 변화율을 나타내며, 다음과 같이 출력 영상 $p^{(t)}$ 의 각 화소에 대한 편미분 집합으로 표현할 수 있다.

식 (7)에서 $\partial L(s, c, p^{(t)}) / \partial p_{i,j,k}^{(t)}$ 는 식 (2)~식 (4)의 정의에 따라, 다음과 같이 나타낼 수 있다.

$$\frac{\partial L(s, c, p^{(t)})}{\partial p_{i,j,k}^{(t)}} = \lambda_{os} \cdot \frac{\partial L_o(c, p^{(t)})}{\partial p_{i,j,k}^{(t)}} + \frac{\partial L_s(s, p^{(t)})}{\partial p_{i,j,k}^{(t)}}, \quad (8)$$

$$\frac{\partial L_o(c, p^{(t)})}{\partial p_{i,j,k}^{(t)}} = \sum_{l \in C} \lambda_o(l) \cdot [\phi_l(c) - \phi_l(p^{(t)})] \circ \frac{\partial \phi_l(p^{(t)})}{\partial p_{i,j,k}^{(t)}}, \quad (9)$$

식 (9)에서 \circ 는 연산자 좌/우에 위치한 두 텐서의 항목별 곱 (element-wise multiplication)을 의미하고, 식 (10)의 $G_{m,n}(\cdot)$ 은 그램 행렬 $G(\cdot)$ 의 m 행 n 열 원소를 말한다. 식 (9)의 $\partial \phi_l(p^{(t)}) / \partial p_{i,j,k}^{(t)}$ 는 한 화소 $p_{i,j,k}^{(t)}$ 의 변화에 대한 스타일 전이 네트워크 출력 $\phi_l()$ 의 변화율을 나타내는 H_l, W_l, C_l 크기의 텐서를 의미하는데, $p_{i,j,k}^{(t)}$ 가 영향을 미치는 일정 크기⁵⁾에서만 값을 가지고 나머지 영역은 모두 0이 되는 특징이 있다. 따라서 텐서 $[\phi_l(c) - \phi_l(p^{(t)})]$ 의 차원이 입력 영상의 차원에 비례함에도 불구하고 식 (9)는 그 차원에 상관없는 양이 되고, 이에 따라 식 (3)에는 계층 별 출력 텐서의 차원에 따른 정규화가 고려되지 않았다. 이와는 대조적으로 식 (10)에서 $G_{m,n}(\cdot)$ 은 그 크기가 $(H_l \times W_l)^2$ 에 비례하고, 그 수가 $1 \leq m, n \leq C_l$ 이므로, 식 (4)에서는 이러한 계층 별 출력 텐서의 차원에 따른 정규화가 고려되었다. 스타일 전이 네트워크에서 이와 같은 정규화가 중요한 이유는, 식 (6)의 각 화소에 대한 갱신 방향은

5) 보다 자세히는, $\phi_l(p^{(n)})$ 에서 $p^{(n)}$ 까지의 역합성곱(Transposed Convolution) 신경망에서 $p_{i,j,k}^{(t)}$ 의 수용장(receptive field) 영역을 의미한다.

식 (9)와 식 (10)의 각 부분 경사도의 가중 합으로 결정되므로 특정 부분 오차가 큰 값을 가지게 되면 해당 오차를 줄이는 방향으로 수렴이 시작되고, 경사 하강법이 국부 최적화 기법임을 고려하면 이러한 초기 수렴 방향이 전체적인 스타일 전이 특성을 좌우할 수 있기 때문이다. 특히, 각 부분 경사도를 정의하는 계층 별 출력 $\phi_l()$ 이 네트워크로 연결되어 이전 계층들의 출력에 영향을 받는 구조일 뿐 아니라, 높은 계층의 출력에 영향을 주는 수용장은 상당히 큰 값을 가지기 때문에(예를 들어, Conv5_4의 수용장 크기는 252x252) 지속적으로 각 계층 별 부분 오차가 줄어드는(혹

은 수렴하는) 것은 어렵기 때문이다. 이와 같은 수렴 특성 해석에 대한 주장을 실험적으로 정당화하기 위해, 그림 3에 식 (4)의 각 부분 스타일 손실⁶⁾에 대한 수렴 과정을 도시하였다.

그림 3은 그림 2.(a)의 1행 4열 스타일 전이 결과를 얻는 수렴 과정에서 각 스타일 부분 손실의 수렴 경향을 도시한 것이다. 그림을 보면, Conv4_4계층에서 정의된 부분 손실이 학습 초기에 가장 빠르게 수렴하는 반면, Conv1_2계층의 부분 손실이 가장 느리게 수렴하는 것을 알 수 있다. 또한, 각 부분 손실의 초기 값이 매우 큰 편차를 보이는 것을

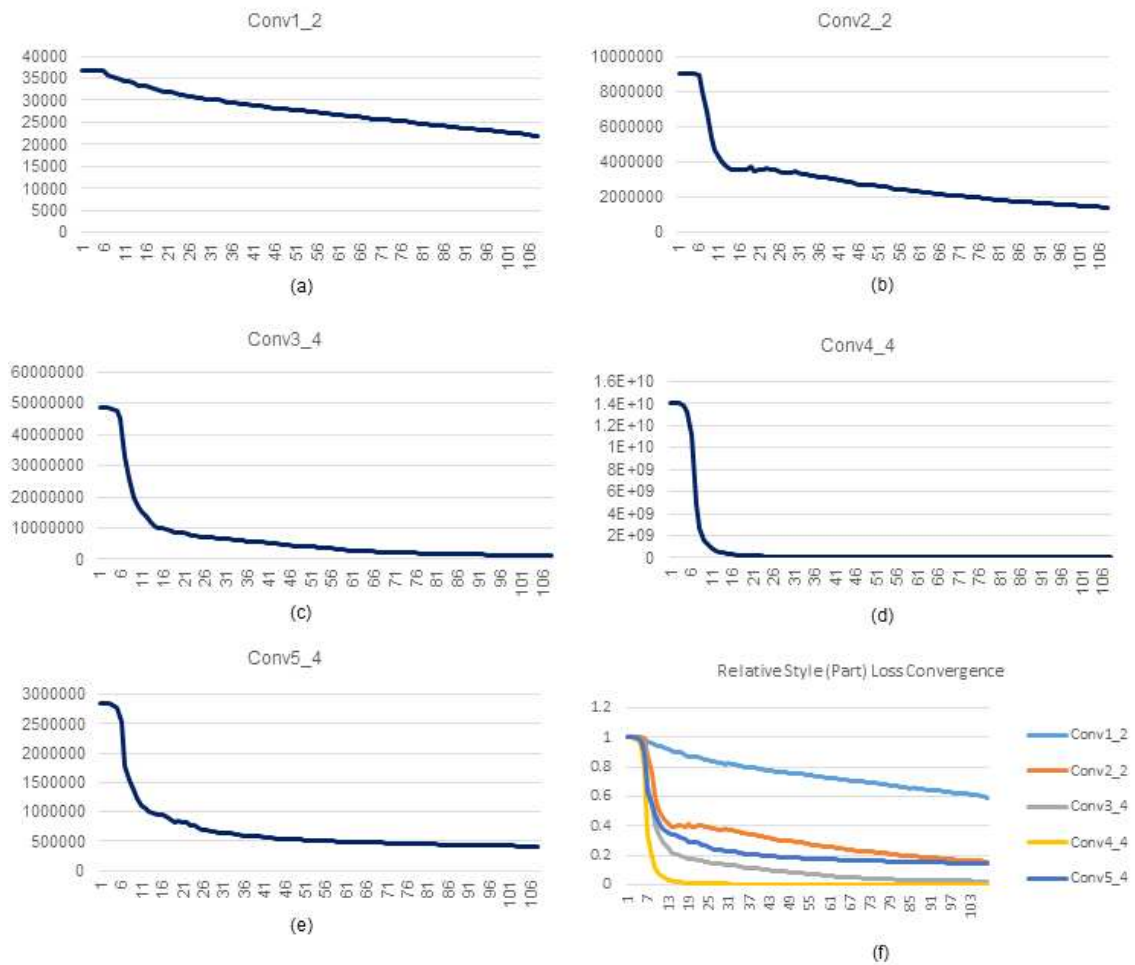


그림 3. 각 스타일 부분 손실의 수렴 경향
 Fig. 3. Convergence trend of each style part loss (defined at each layer)

6) 식(4)를 $L_s(s, x) = \sum_{j \in S} \lambda_s(j) \cdot L_s(s, x, j)$ 라 표현했을 때, $L_s(s, x, j)$ 를 계층 j에서 정의된 부분 스타일 손실 함수라 한다.

알 수 있는데, 그림 3.(a)의 초기 부분 손실을 1이라 할 때, 그림 3.(b)-(e)에 도시된 각 부분 손실의 초기 값은 245.9, 1317.6, 381620.0, 그리고 77.33에 해당하고, 따라서 각 부분 손실을 하나의 그래프에서 비교하기가 매우 어렵다. 그림 3.(f)는 각 부분 손실의 초기치를 모두 1로 정규화 한 후, (a)-(e)를 하나의 그래프에 도시한 것으로 각 부분 손실의 초기 수렴 경향을 살펴볼 수 있다. 이러한 스타일 부분 손실 수렴 경향 그래프에서 확인할 수 있는 사실은 Conv4_4계층에서 정의된 초기 부분 손실이 다른 계층의 초기 부분 손실에 비해 최소 300배 이상 (Conv3_4 대비) 크기 때문에, 식 (10)의 스타일 오차 경사도 계산에서 해당 계층의 부분 경사도가 다른 계층으로부터의 부분 경사도를 크게 압도함으로써 수렴 과정에서 Conv4_4계층의 부분 오차가 빠르게 감소할 뿐 아니라, 100배 차이의 부분 손실 가중치 조절을 통해 스타일 변환을 수행했던 그림 2.(b)에 대한 결과를 이해할 수 있도록 해준다.

2. 개선된 제안 스타일 전이 방법

앞 절의 스타일 전이 네트워크 수렴 특성은 그림 1의 VGG-19 네트워크가 정규화 레이어 (예를 들어, Batch Normalization [21] 또는 Instance Normalization [14] 등) 없이 학습되어, 각 계층별 출력 텐서의 크기가 큰 차이를 보이기 때문이다.

그림 4에 이와 같은 계층별 출력 파워의 차이를 예시하였다. 그림은 ‘Starry Night’ 영상(그림 5.(b) 참조)을 입력하여

구한 VGG-19 네트워크의 각 계층 별 출력 크기를 도시한 것으로, 4_1, 4_2 계층의 출력이 다른 계층, 특히 계층 1이나 계층 5에 비해 매우 큰 값을 가지는 것을 볼 수 있다. 이러한 계층 별 출력 크기의 차이로부터 식 (4)의 부분 스타일 손실의 큰 차이가 유발되고, 이러한 차이가 그림 3에 도시한 바와 같이 각 계층 별 수렴 특징을 결정하게 되며, 이에 따라 그림 2의 부분 스타일 손실에 대한 가중치 변화에 대한 수렴 결과 영상의 제한적 특징이 나타나는 것으로 볼 수 있다. 뿐만 아니라, 이러한 계층 별 출력의 크기 차이는 입력 영상에 따라서도 크게 달라지기 때문에, 그림 2의 실험 결과와 같이 적절한 스타일 전이를 위한 초-매개 변수의 설정에 큰 차이를 요구할 수 있다.

이러한 스타일 전이 네트워크의 계층 별 출력 크기 차이에 기인한 스타일 전이 결과의 불안정성을 제거하고, 직관적 초-매개 변수 설정을 통한 스타일 전이 특성을 안정적으로 조절하기 위해, 본 논문에서는, 식 (3), 식 (4)의 오브제 손실과 스타일 손실을 식 (11), 식 (12)와 같이 정규화 된 형태로 수정하여 사용하는 방법을 제안한다.

$$L_o(o, x) = w \cdot \sum_{j \in C} \lambda_o^R(j) \cdot \|\phi_j(o) - \phi_j(x)\|_2^2 / E_j^o, \quad (11)$$

$$L_s(s, x) = w \cdot \sum_{j \in S} \lambda_s^R(j) \cdot \|G(\phi_j(s)) - G(\phi_j(x))\|_F^2 / E_j^G \quad (12)$$

여기서, E_j^o 와 E_j^G 는 오브제 영상의 j-번째 계층 출력 에너지와 스타일 영상의 j-번째 계층 출력으로 구성된 Gram

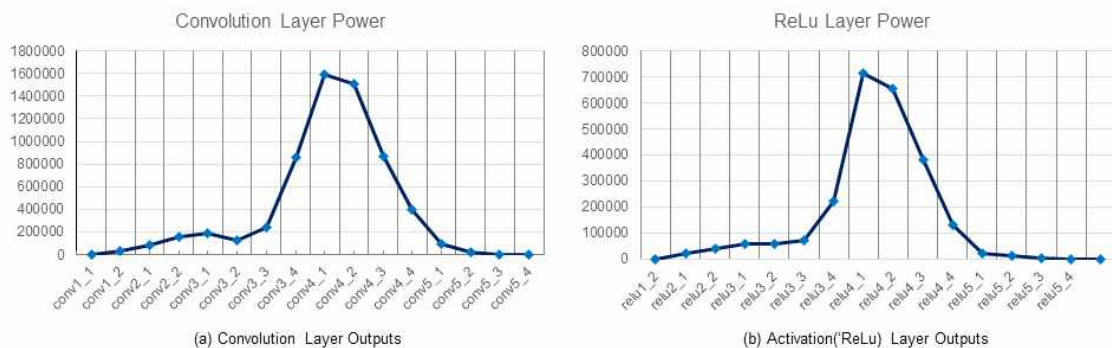


그림 4. VGG-19 [17] 네트워크 계층 별 출력 파워 (‘Starry Night’ 영상 입력)

Fig. 4. Output Power for each Layer of VGG-19 Network [17] (based on the ‘Starry Night’ image)

행렬의 에너지를 의미하는 것으로, 구체적으로 식 (13)과 같이 정의되며, 입력된 오브제/스타일 영상의 해당 계층 출력으로부터 구성되는 부분 손실이 전체 손실에서 균일한 영향을 가질 수 있도록 고려한 것이다.

$$E_j^o = \sum_{H_j, W_j, C_j} \phi_j^2(o), \quad E_j^G = ||G(\phi_j(s))||_F^2 \quad (13)$$

또한, 식 (11), 식 (12)의 부분 손실 가중치 $\lambda_o^R(j), \lambda_s^R(j)$ 도 식 (14)와 같이 정규화 함으로써, 오브제 손실과 스타일 손실의 각 계층 별 가중치 차이가 상대적인 관계만 가지도록 하고, 그 총 합이 1이 되도록 함으로써, 식 (2)의 λ_{os} 가 정규화 된 일정 크기 손실에 대한 상대적인 가중치를 의미할 수 있도록 하였다. 마지막으로, 식 (11), 식 (12)의 w 는 오브제 손실과 스타일 손실의 크기를 적당한 크기로 조절하기 위한 명목적 가중치로, 본 논문에서는, 입력된 오브제 영상의 화소수를 그 값으로 사용하였다.

$$\lambda_o^R(j) = \lambda_o(j) / \sum_{k \in C} \lambda_o(k), \quad \lambda_s^R(j) = \lambda_s(j) / \sum_{k \in C} \lambda_s(k) \quad (14)$$

이와 같은 정규화 방식은 입력 영상의 특성이 반영된 부분 손실의 정규화, 오브제 손실과 스타일 손실의 크기 균일화 및 입력 영상의 해상도에 따른 부분 손실의 상대적 비율 안정화를 고려한 것으로, 입력 영상의 특징 및 해상도에 따른 초-매개 변수 선택의 어려움과 제한적 스타일 전이 결과에 대한 문제(그림 2 참조)를 해결하고, 직관적 파라미터 선택으로 보다 안정적이고 다양한 스타일 전이 결과를 제공할 수 있다.

IV. 실험 및 결과

제안된 정규화 방식이 제공하는 스타일 전이의 우수성을 검증하기 위해, 그림 5의 실험 영상을 대상으로 기존 방식 [5]과 제안 방법을 이용한 스타일 전이 결과를 비교하였다. 스타일 전이를 위한 식 (1)의 최적화는 Tensorflow [22]에서 제공하는 L-BFGS [23] 최적화기를 통해 수행되었고, Titan-V GPU를 장착한 Ubuntu 18.04 시스템이 사용되었다. 식 (1)의 x 를 위한 초기 영상은 입력된 오브제 영상 o 를 사용하였으며, 최적화를 위한 반복 회수는 600회로 설정하였다.

또한, 식 (3)과 식 (11)의 오브제 손실 $L_o(\cdot, \cdot)$,의 정의를 위한 계층 집합은 $C = \{Conv4_2\}$ 를 사용하였고, 식 (4)와 식 (12)의 스타일 손실 $L_s(\cdot, \cdot)$,을 정의하기 위한 계층 집합은 $S = \{Conv1_2, Conv2_2, Conv3_4, Conv4_4, Conv5_4\}$ 를 사용하였다.

그림 5.(a)의 오브제 영상과 그림 5.(b)-(d)의 스타일 영상을 대상으로 다양한 초-매개 변수 및 오브제 영상의 입력 해상도 변화에 따른 기존 방식의 스타일 전이 결과를 그림 2에 도시하였다. 앞서 설명한 바와 같이, 기존 방식에서는 입력된 스타일 영상에 따라 적절한 스타일 전이를 제공할 수 있는 식 (2)의 가중치 λ_{os} 값이 크게 달라질 뿐 아니라, 계층 별 스타일 부분 손실 가중치를 의미하는 식 (4)의 $\lambda_s(j)$ 변화에 따른 스타일 전이 결과의 다양성이 매우 제한적이고, 입력 오브제 영상의 해상도에 따라 동일 초-매개 변수가 제공하는 스타일 전이 결과에 큰 차이를 보이는 문제가 있다.

반면, 본 논문에서 제안하는 정규화 방식 (식 (11) ~ 식



그림 5. 성능 비교를 위한 실험 영상
Fig. 5. Test Images for the Comparison of Style Transfer Results



(a) Results of the Proposed Style Transfer with the changes of λ_{os}

$$\lambda_{os}^2(j) = \begin{bmatrix} 100 & 1 & 1 & 1 & 1 \\ 104 & 104 & 104 & 104 & 104 \end{bmatrix} \lambda_{os}^2(j) = \begin{bmatrix} 1 & 100 & 1 & 1 & 1 \\ 104 & 104 & 104 & 104 & 104 \end{bmatrix} \lambda_{os}^2(j) = \begin{bmatrix} 1 & 1 & 100 & 1 & 1 \\ 104 & 104 & 104 & 104 & 104 \end{bmatrix} \lambda_{os}^2(j) = \begin{bmatrix} 1 & 1 & 1 & 100 & 1 \\ 104 & 104 & 104 & 104 & 104 \end{bmatrix} \lambda_{os}^2(j) = \begin{bmatrix} 1 & 1 & 1 & 1 & 100 \\ 104 & 104 & 104 & 104 & 104 \end{bmatrix}$$



(b) Results of the Proposed Style Transfer with the changes of $\lambda_{os}^2(j)$



(c) Results of the Proposed Style Transfer with the changes of Image Size

그림 6. 초-매개변수 변화에 따른 제안 스타일 전이 기법 실험 결과
 Fig. 6. Simulation results of the proposed style transfer with the changes of hyper-parameters

(14) 참고)을 사용하는 경우의 스타일 전이 결과는 그림 6과 같다. 우선, 그림 6.(a)의 결과를 보면, $10^{-2} \leq \lambda_{os} \leq 10^{-3}$ 의 초-매개 변수 범위에서 모든 스타일 입력 영상에 대해 그림 2.(a)의 적절한 스타일 전이 수준과 유사한 기준의 스타일 전이가 가능한 것을 알 수 있다. 이러한 결과는, 기존 방식의 경우(그림 2.(a)), 서로 다른 스타일 영상으로부터 적절한 수준의 스타일 전이를 얻기 위해 최대 100배 차이가 나는 초-매개변수 λ_{os} 의 설정이 필요했던 것과는 대조적으로, 동일한 초-매개 변수 값의 설정으로 동일 수준의 스타일 전이 강도를 얻을 수 있음을 나타낸다. 두 번째로, 그림 2.(b)와 그림 6.(b)의 결과)를 비교하면, 제안 방식에서는 스타일 손실의 계층 별 손실 가중치(식 (12)의 $\lambda_s^R(j)$) 변화에 따라 차별적인 스타일 전이 결과를 얻을 수 있음을 확인할 수 있다. 특히, 오브제 영상의 하늘 영역이나 강물 영역과 같이 복잡한 오브제가 위치하지 않는 빈 공간의 스타일 전이 결과를 보면, 점차 높은 계층의 가중치가 강화되면서 (1열의 결과에서 5열의 결과 순으로) 더 넓은 수용장을 가지는 계층의 특징 벡터로부터 정의되는 그램-행렬의 매칭이 강조되어, 더 넓은 스타일 영상 영역에 존재하는 높은 수준의 구조적 특징이 스타일로 반영되는 것을 관찰할 수 있다. 마지막으로, 그림 6.(c)의 오브제 입력 영상 해상도에 따른 스타일 전이 결과를 살펴보면, 그림 2.(c)의 기존 방식이 제공하던 결과에 비해, 완벽하진 않지만, 스타일 전이 특징이 조금 더 안정화 된 것을 볼 수 있다. 낮은 해상도에서 관찰되는 큰 덩어리 형태의 스타일 전이 결과가 해상도가 커지면서 점차 세밀한 작은 덩어리 형태의 스타일 전이 결과로 변화하는 결과는 VGG-19 네트워크의 계층 별로 고정된 수용장 크기에 의한 현상으로, 해상도가 큰 영상의 경우 해당 수용장 크기가 전체 영상에서 작은 영역에 해당되기 때문이다. 이러한 문제는, [7]에서 지적된 바와 같이, 최선의 스타일 전이 상황(sweet spot)에 해당하는 특정 해상도 스타일 전이 결과를 점차 확대하며, 식 (1)의 x 를 위한 초기 영상으로 입력하여 재귀적으로 스타일 전이를 수행함으로써 해결할 수 있다. 다만, 그림 2.(c)와 그림 6.(c)의 결

과가 의미하는 바는, 식 (3)과 식 (4)의 기존 방식 손실 정의에서는 입력 해상도에 따라 그 손실 값의 상대적 크기가 변화하여 해상도에 따른 적절한 λ_{os} 를 재설정해야 하는 어려움이 제안된 정규화 기반 스타일 전이 기법에서는 해결되었기 때문에, 재귀적 스타일 전이를 통해 해상도에 따른 스타일 전이 강도를 쉽게 정규화 할 수 있게 되었다는 것이다.

이상에서 설명한 제안 방식의 스타일 전이 우수성을 보다 객관적으로 검증하기 위해, 10명의 예술전공 학생을 대상으로 5점 척도의 주관적 스타일 전이 적절성 검증 실험을 수행하였다. 실험 참가자들에게는, 우선, 스타일 전이의 형태적 변형 특성을 설명하면서 그림 2의 영상들 중 그림 5.(c)의 스타일 영상을 사용한 스타일 전이 영상들을 보여주었다. 그 후, 그림 2와 그림 6의 나머지 스타일 전이 영상들⁸⁾ 대상으로 스타일 전이가 잘 되었다고 느끼는 정도를 5점 척도로 조사하였다. 조사를 통해, 실험 참가자가 느끼는 기존 방식(그림 2)과 제안 방식(그림 6)에 대한 상대적 비교 결과를 얻기 위해, 각 평가는 그림 2와 그림 6에서 서로 대응되는 두 영상을 무작위 순서로 5초씩 보여주고, 그 후 5초간 두 영상에 대한 점수를 매기도록 하는 형식으로 수행되었다. 이러한 주관적 스타일 전이 평가 실험의 결과를 표 1에 정리하였다. 각 결과는 10명의 실험 참가자들이 제공한 각 스타일 전이 영상에 대한 주관적 평가 결과를 평균±표준편차의 형식으로 표현한 것이며, 각 결과의 위치는 그림 2와 그림 6에 도시된 그림의 위치에 대응된다.

우선, 그림 2.(a)와 그림 6.(a)의 비교 실험 결과를 살펴보면, 평균 4.0(‘스타일 전이가 잘 이루어졌다’) 이상의 고수준 스타일 전이 결과를 제공할 수 있는 초-매개 변수 λ_{os} 의 범위(음영 표시된 결과 영역)가 기존 방식의 경우에는 입력 영상에 따라 크게 차이가 나는 반면, 제안 방식에서는 동일한 특정 범위에서 제공되는 것을 볼 수 있다. 이러한 결과는 앞서 설명한 그림 6의 결과 분석을 정당화 한다. 두 번째로, 그림 2.(b)와 그림 6.(b)의 비교 실험 결과 분석을 위해, 평균 2.0(‘과도하거나 매우 약하게 스타일 전이가 이루어졌

7) 그림 6.(b)의 결과 생성을 위한 λ_{os} 파라미터는 모든 스타일 영상에 대해 10^{-3} 의 값을 사용하였다.

8) 구체적으로, 그림 2.(a)와 그림 6.(a)의 첫 두 행에 위치한 영상들, 그림 2.(b)와 그림 6.(b)의 첫 두 행에 위치한 영상들, 그리고 그림 2.(c)와 그림 6.(c)의 영상들을 의미한다.

표 1. 스타일 전이 적절성에 대한 주관적 평가 실험 결과

Table. 1. Subjective evaluation results on the appropriateness of style transfer

evaluation results for the images of Fig 2.(a) and Fig 6.(a)						
1 st row im- ages	Fig 2.(a)	1.0±0.00	1.3±0.48	2.3±0.48	4.3±0.48	4.9±0.30
	Fig 6.(a)	1.0±0.00	1.8±0.40	4.2±0.40	4.8±0.40	5.0±0.00
2 nd row im- ages	Fig 2.(a)	1.0±0.00	4.3±0.48	4.6±0.52	3.6±0.52	3.5±0.53
	Fif 6.(a)	1.0±0.00	2.2±0.40	4.2±0.40	4.9±0.30	4.2±0.40
evaluation results for the images of Fig 2.(b) and Fig 6.(b)						
1 st row im- ages	Fig 2.(b)	1.8±0.42	1.9±0.36	2.0±0.00	4.2±0.42	2.1±0.31
	Fig 6.(b)	2.5±0.50	2.9±0.30	2.8±0.40	4.5±0.50	4.3±0.46
2 nd row im- ages	Fig 2.(b)	1.3±0.48	1.0±0.00	1.1±0.32	4.5±0.53	1.2±0.42
	Fif 6.(b)	2.8±0.40	4.2±0.40	4.9±0.30	4.9±0.30	4.2±0.40
evaluation results for the images of Fig 2.(c) and Fig 6.(c)						
1 st row im- ages	Fig 2.(c)	2.0±0.00		4.5±0.53		1.4±0.52
	Fig 6.(c)	3.3±0.48		4.7±0.48		2.1±0.32
2 nd row im- ages	Fig 2.(c)	2.0±0.00		4.5±0.53		1.0±0.00
	Fif 6.(c)	2.1±0.32		4.8±0.42		2.3±0.48

다) 이상의 영역을 음영으로 표시하였다. 제안 방식은 각 레이어의 가중치 강화 결과가 모두 평균 2.0 이상의 수준을 제공하고 있지만, 기존 방식의 경우에는 많은 영역에서 스타일 전이가 이루어졌다고 느낄 수 있는 수준(평균 2.0)에 미치지 못한다는 결과를 확인할 수 있다. 뿐만 아니라, 제안 방식의 경우에는 평균 4.0 이상의 고수준 스타일 전이로 인식되는 구간이 상대적으로 매우 넓게 분포할 뿐 아니라, 다른 구간에서도 평균 3.0(‘다소 약하거나, 조금 과장되게 스타일 전이가 이루어졌다’)에 보다 가까운 결과를 제공하고 있다. 낮은 레이어의 가중치 강화가(그림 2.(b), 그림 6.(b)의 앞쪽 열에 위치한 결과 영상들), 낮은 수준의(작은 수용장에서 분석된 영상의 특징) 특징이 강조되는 스타일 전이 결과를 제공한다는 특성을 고려하면, 그림 6.(b)의 처음 1~3열에서 이와 같은 결과가 도출된 것은 제안 방식이 레이어 별 가중치 조합을 통해 매우 다양한 스타일 전이 결과를 도출할 수 있다는 주장을 정당화하는 것이다. 마지막으로, 그림 2.(c)와 그림 6.(c)의 주관적 비교 실험 결과를 살펴보면, 512 해상도(가운데 결과)를 제외한 결과에서 제안 방식이 기존 방식을 크게 개선한 것을 확인할 수 있다. 비록 그 개선 정도가 평균 4.0 수준에 이르지 못하는 제한적인 수준이긴 하지만, 1024 해상도(오른쪽 결과)에서 기존 결과

가 제공하던 평균 1.0(‘스타일 전이가 이루어지지 않았다’) 수준의 결과를 모두 2.0 이상의 수준으로 개선한 것과, 256 해상도의 첫 번째 행 영상에 대한 결과를 2.0에서 3.3 수준으로 개선한 것은, 인식되는 스타일 전이 결과의 수준 자체를 바꾼, 매우 고무적이고 유의미한 결과라 할 수 있다.

V. 결론

신경망 기반 스타일 전이 기법은 영상의 고차원적 구조적 특징을 적절하게 반영하여 높은 수준의 전이 품질을 제공함으로써 많은 관심을 모으고 있다. 이러한 관심은 스타일 전이에 대한 다양한 최신 연구 결과들이 소프트웨어 형태로 일반에 공개됨으로써 더욱 가속화되고 있으며, NVIDIA^[24]나 Google^[25]과 같은 회사들도 이러한 일반 공개에 동참하면서 다양한 활용 결과들이 나타나고 있다.

본 논문에서는 가장 높은 스타일 전이 품질과 다양성을 제공할 수 있는 기술 중 하나로 알려진 Gatys 등의 획기적인 방식^[5]을 대상으로, 직관적인 초-매개변수 설정을 통해 보다 안정적이고 더욱 다양한 스타일 전이 결과가 생성될 수 있도록 손실함수의 정규화 방법을 제안하였다. 제안된

방식은 입력 영상의 특성이 반영된 부분 손실의 정규화, 오브제 손실과 스타일 손실의 상대적 크기 균일화 및 입력 영상의 해상도에 따른 부분 손실의 상대적 비율 안정화를 고려한 조치로, 대상 방식이 가지고 있던 입력 영상의 특징 및 해상도에 따른 초-매개 변수 선택의 어려움과 제한적 스타일 전이 결과에 대한 문제를 해결하기 위한 것이다. 제안 방식의 우수성을 검증하기 위해, 본 논문에서는, 광범위한 초-매개변수 설정에 따른 스타일 전이 실험 결과를 비교하였으며, 이를 통해 제안 방식이 직관적 파라미터 선택으로 보다 안정적이고 더욱 다양한 스타일 전이 결과를 제공할 수 있음을 보였다.

이러한 제안 방식은 다양한 분야의 아티스트들과 협업을 통해 이미 여러 차례 전시에 활용된 바 있으며^[26], 초-고해상도 영상에 대한 스타일 전이 및 비디오, VR(Virtual Reality) 미디어의 스타일 전이 등으로 전이 대상 미디어를 확장함으로써 더욱 많은 활용이 기대된다.

참 고 문 헌 (References)

- [1] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," *Proc. of the IEEE Int. Conf. on Computer Vision(ICCV)*, 20-25 Sept., Kerkyra, Greece, pp.1033-1038, 1999..
- [2] A. Efros and W. Freeman, "Image quilting for texture synthesis and transfer," *Proc. of the SIGGRAPH*, 12-17 Aug., Los Angeles, U.S.A., pp.341-346, 2001.
- [3] D. Heeger and J. Bergen, "Pyramid-based texture analysis/synthesis," *Proc. of the SIGGRAPH*, 06-11 Aug., Los Angeles, U.S.A., pp.229-238, 1995.
- [4] O. Frigo, N. Sabater, J. Delon, and P. Hellier, "Spit and match: example-based adaptive patch sampling for unsupervised style transfer," *Proc. of the IEEE Conf. on Computer Vis. and Patt. Recog.(CVPR)*, 26 June-01 July, Las Vegas, U.S.A., 2016.
- [5] L. Gatys, A. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Image style transfer using convolutional neural networks," *Proc. of the IEEE Conf. on Computer Vis. and Patt. Recog.(CVPR)*, 26 June-01 July, Las Vegas, U.S.A., 2016.
- [6] Y. Jing, Y. Liu, Y. Yang, Z. Feng, Y. Yu, D. Tao, and M. Song, "Stroke controllable fast style transfer with adaptive receptive fields," *Proc. of the European Conf. on Computer Vision(ECCV)*, 08-14 Sept., Munich, Germany, 2018.
- [7] L. Gatys, A. Ecker, M. Bethge, A. Hertzman, and E. Shechtman, "Controlling perceptual factors in neural style transfer," *Proc. of the IEEE Conf. on Computer Vis. and Patt. Recog.(CVPR)*, 22-25 July, Honolulu, U.S.A., 2017.
- [8] A. Gupta, J. Johnson, A. Alahi, and L. Fei-Fei, "Characterizing and improving stability in neural style transfer," *Proc. of the IEEE Int. Conf. on Computer Vision(ICCV)*, 22-29 Oct., Venice, Italy, 2017.
- [9] H. Huang, H. Wang, W. Luo, L. Ma, W. Jiang, X. Zhu, Z. Li, and W. Liu, "Real-time neural style transfer for videos," *Proc. of the IEEE Conf. on Computer Vis. and Patt. Recog.(CVPR)*, 22-25 July, Honolulu, U.S.A., 2017.
- [10] M. Ruder, A. Dosovitskiy, and T. Brox, "Artistic style transfer for videos and spherical images," *Int. J. of Computer Vision*, vol.126, no.11, pp.1199-1219, 2018.
- [11] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *Proc. of the European Conf. on Computer Vision(ECCV)*, 8-16 Oct., Amsterdam, Netherland, 2016.
- [12] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis," *Proc. of the IEEE Conf. on Computer Vis. and Patt. Recog.(CVPR)*, 22-25 July, Honolulu, U.S.A., 2017.
- [13] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," *Proc. of the Int. Conf. on Learning Representations*, 24-26 Apr., Paris France, 2017.
- [14] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," *Proc. of the IEEE Int. Conf. on Computer Vision(ICCV)*, 22-29 Oct., Venice, Italy, 2017.
- [15] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M. Yang, "Universal style transfer via feature transforms," *Proc. of the Conf. on Neural Information Processing Systems (NIPS)*, 04-09 Dec., Long Beach U.S.A., 2017.
- [16] L. Sheng, Z. Lin, J. Shao, and X. Wang, "Avatar-net: multi-scale zero-shot style transfer by feature decoration," *Proc. of the IEEE Conf. on Computer Vis. and Patt. Recog.(CVPR)*, 18-22 June, Salt Lake City U.S.A., 2018.
- [17] K. Simonyan and A. Zissermann, "Very deep convolutional networks for large-scale image recognition," *Proc. of the Int. Conf. on Learning and Representations(ICLR)*, 7-9 May, San Diego U.S.A., 2015.
- [18] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol.1, no.4, pp.541-551, Dec. 1989.
- [19] J. Yosinski, J. Clune, . Anh Nguyen, T. Fucht, and H. Lipson, "Understanding neural networks through deep visualization," *arXiv:1506.06579*, 2015, <http://arxiv.org/abs/1506.06579> (accessed Nov. 18, 2020)
- [20] W. Samek, A. Binder, G. Montavon, S. Bach, and K.-R. Muller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Trans. on Neural Networks and Learning Systems*, vol.28, no.11, pp.2660-2673, Nov. 2017.
- [21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proc. Int. Conf. on Machine Learning*, 06-11 July, Lille, France, 2015.
- [22] M. Abadi, et. al., "Tensorflow: A system for large-scale machine learning," *Proc. OSDI'16*, Nov.2-4, GA, U.S.A., 2016.
- [23] D. Liu and J. Nocedal, "On the limited memory BFGS method for large

scale optimization,” *Mathematical Programming*, vol.45, Issue1-3, pp.503-528, Springer 1989.

- [24] Fast Photo Style, <https://github.com/NVIDIA/FastPhotoStyle> (accessed Nov. 18, 2020)
- [25] neural style transfer using tf.keras, https://www.tensorflow.org/tutorials/generative/style_transfer (accessed Nov. 18, 2020)

- [26] KOCCA, *The excavation of AI based creative atelier and the research and development on creating technology*, <https://doi.org/10.23000/TRKO201900002287> (accessed Nov. 18, 2020)

저 자 소 개



최 인 성

- 2008년 : 서울과학기술대학교 제어계측과 공학사
- 2004년 ~ 현재 : (주)프로메가바이오시스템스 R&D 차장
- 2017년 ~ 현재 : SMIT 인공지능응용소프트웨어학과 재학
- ORCID : <https://orcid.org/0000-0002-5054-2524>
- 주관심분야 : 컴퓨터비전, 딥러닝/머신러닝 기반의 머신비전 솔루션



김 용 구

- 1993년 : 연세대학교 전기공학과 공학사
- 1995년 : 연세대학교 전기및컴퓨터공학과 공학석사
- 2001년 : 연세대학교 전기전자공학과 공학박사
- 2002년 ~ 2006년 : ㈜온타임텍 멀티미디어연구소 연구소장/이사
- 2009년 ~ 현재 : SMIT 뉴미디어학부 교수
- ORCID : <http://orcid.org/0000-0002-8905-1984>
- 주관심분야 : 초실감미디어, 컴퓨터비전, 딥-러닝, 비디오 압축, 멀티미디어 시스템