JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 28 No. 8, pp. 31-38, August 2023
https://doi.org/10.9708/jksci.2023.28.08.031

Optimization of attention map based model for improving the usability of style transfer techniques

Junghye Min*

*Professor, Dept. of Computer Science, Inha Technical College, Incheon, Korea

[Abstract]

Style transfer is one of deep learning-based image processing techniques that has been actively researched recently. These research efforts have led to significant improvements in the quality of result images. Style transfer is a technology that takes a content image and a style image as inputs and generates a transformed result image by applying the characteristics of the style image to the content image. It is becoming increasingly important in exploiting the diversity of digital content. To improve the usability of style transfer technology, ensuring stable performance is crucial. Recently, in the field of natural language processing, the concept of Transformers has been actively utilized. Attention maps, which forms the basis of Transformers, is also being actively applied and researched in the development of style transfer techniques. In this paper, we analyze the representative techniques SANet and AdaAttN and propose a novel attention map-based structure which can generate improved style transfer results. The results demonstrate that the proposed technique effectively preserves the structure of the content image while applying the characteristics of the style image.

▸ **Key words**: Style Transfer, Self Attention, Feed-forward network, Image Processing, Deep Learning

[요 약]

딥러닝 기반 영상 처리 기술 중 최근 활발히 연구되어 많은 성능 향상을 이룬 기술 중 하나는 스타일 전이 (Style Transfer) 기술이다. 스타일 전이 기술은 콘텐츠 영상과 스타일 영상을 입력받아 콘텐츠 영상의 스타일을 변환한 결과 영상을 생성하는 기술로 디지털 콘텐츠의 다양성을 확보하는데 활용할 수 있어 중요성이 커지고 있다. 이런 스타일 전이 기술의 사용성을 향상하기 위해서는 안정적인 성능의 확보가 중요하다. 최근 자연어 처리 분야에서 트랜스포머 (Transformer) 개념이 적극적으로 활용됨에 트랜스포머의 기반이 되는 어텐션 맵이 스타일 전이 기술 개발에도 활발하게 적용되어 연구되고 있다. 본 논문에서는 그중 대표가 되는 SANet과 AdaAttN 기술을 분석하고 향상된 스타일 전이 결과를 생성 할 수 있는 새로운 어텐션 맵 기반 구조를 제안한다. 결과 영상은 제안하는 기술이 콘텐츠 영상의 구조를 보존하면서도 스타일 영상의 특징을 효과적으로 적용하고 있음을 보여준다.

▸ **주제어**: 스타일 전이, 셀프 어텐션, 피드포워드 네트워크, 영상 처리, 딥러닝

• First Author: Junghye Min, Corresponding Author: Junghye Min
*Junghye Min (jhmin@inhatc.ac.kr), Dept. of Computer Science, Inha Technical College
• Received: 2023. 07. 12, Revised: 2023. 07. 26, Accepted: 2023. 08. 01.

Copyright © 2023 The Korea Society of Computer and Information
http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

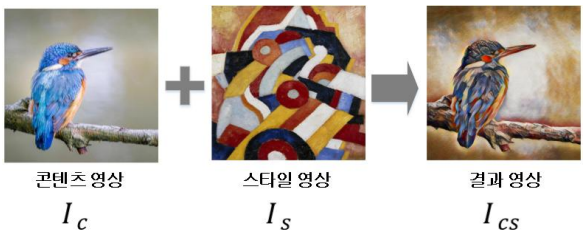

Fig. 1. Style Transfer

딥러닝 기반 영상 처리 기술 중 최근 활발히 연구되어 많은 성능 향상을 이룬 기술 중 하나는 스타일 전이 (Style Transfer) 기술이다. 스타일 전이 기술은 Fig.1 에서 볼 수 있듯이 콘텐츠 영상과 스타일 영상을 입력받아 콘텐츠 영상의 스타일을 변환한 결과 영상을 생성한다.

이러한 스타일 전이 기술은 영상 편집, 애니메이션 제작, 가상 현실 콘텐츠 제작 등 적용할 수 있는 분야가 다양하다. 특히 최근에 디지털 아트 시장의 커지고 있어 영상에 예술적 효과를 적용할 수 있는 기술인 스타일 전이 기술의 중요성이 커지고 있다.

스타일 전이 기술의 사용성 향상을 위해서는 두 가지 기준이 충족되어야 한다. 첫째는 콘텐츠 영상의 구조와 정보가 결과 영상에 반영되어야 한다는 것이다. 두 번째 기준은 스타일 영상의 스타일 특징이 결과 영상에 반영되어야 한다는 것이다. 콘텐츠 영상의 구조와 디테일을 보존하려는 노력은 스타일 영상의 특징을 반영하지 못하게 하는 결과를 가져올 수 있다. 반면에 스타일을 강조하는 과정에서 콘텐츠 영상의 구조 정보를 잃어버릴 수 있어 두 가지 기준을 동시에 만족하게 하는 것은 어려운 주제이다.

본 논문의 목적은 이 두 가지 기준을 균형을 맞춰 만족시킬 수 있는 기술을 제안하는 것이다. 어텐션 맵은 자연어 처리 분야뿐만 아니라 영상 처리 분야에서 전반적으로 활발히 사용되고 있다. 본 논문에서는 어텐션 맵을 스타일 전이에 성공적으로 적용한 두 기술 SANet과 AdaAttN의 구조를 분석하고 성능 향상을 위해 최적화된 구조를 제안한다. 2장에서는 관련 기술의 연구 동향을 소개한다. 3장에서는 어텐션 기반 스타일 전이 기술인 SANet과 AdaAttN의 구조와 한계를 분석한다. 4장에서는 새로운 어텐션 맵 기반 프레임워크를 제안한다. 제안한 기술을 포함한 다양한 모델을 이용하여 생성된 결과 영상의 비교와 실험 방법에 대한 정보는 5장에서 소개한다. 6장에선 제안한 기술의 의미와 앞으로의 연구 방향을 소개한다.

II. Preliminaries

1. Relater works

L.A.Gatys[1]가 CNN (Convolutional Neural Network)에서 추출한 영상의 특성을 스타일 전이에 효과적으로 적용한 결과를 발표한 이후로 스타일 전이 기술은 활발히 연구되고 많은 성능 향상을 이루었다. L. A. Gatys는 CNN 중에서도 영상 인식에 사용되는 모델인 VGG Network[3]을 이용하여 콘텐츠 영상과 스타일 영상의 특성을 획득하였다. 초기 스타일 전이 방법들은 CNN을 영상 특성을 추출하는 방법으로만 사용하고 결과 영상은 손실 함수의 최적화 (optimization) 과정을 통해서 생성되었다. 이러한 최적화 과정은 반복에 의한 연산량이 문제가 있어 그 대안으로 CNN을 영상 생성 과정에서도 사용하는 방법이 연구되었다.

피드 포워드 네트워크 (Feed-forward Network)를 결과 영상 생성단계에 적용되기 시작한 시기에는 하나의 네트워크를 특정 영상의 스타일 적용에만 사용할 수 있는 방법이 제안되었다 [2]. 이후에는 임의의 스타일 이미지를 사용해서도 결과 영상을 생성 할 수 있는 네트워크가 제안되었다. ADAIN [5] 은 콘텐츠 영상의 색상 분포를 스타일 영상의 분포에 맞게 변형하는 방법을 제안하였다. 이 기술은 두 영상의 평균과 분산을 맞춰주는 과정을 포함한다. WCT[4] 는 ADAIN 기술에서 제안한 평균과 분산을 맞춰주는 과정이 공분산 (Covariance)까지 고려하도록 설계되어 있고 화이트닝 (whitening)과 컬러링 (coloring), 두 단계에 걸쳐서 수행된다. Avatar-Net[7]은 스타일 교환 (style swap) 기술과 ADAIN 기술을 결합한 다중 스케일 프레임 워크를 제안하였다.

자연어 처리 분야에서 트랜스포머[9] 기술이 활발하게 사용됨에 따라 트랜스포머의 기본인 어텐션 개념 [9]을 활용한 방법들이 제안되었다. Y.Deng [10]은 셀프 어텐션 개념을 콘텐츠 영상과 스타일 영상의 특성을 추출하는데 각각 적용한 후 결합하는 방법으로 영상을 생성하였다. AAMS [8]는 셀프 어텐션과 다중 스트로크 적용 기술을 결합하여 영상을 생성하였다. SANet [6]은 콘텐츠 영상의 피처 맵과 스타일 영상의 피처 맵 간의 상관관계를 도출하기 위하여 어텐션 맵을 사용하였다. AdaAttN [11]은 하위 레벨 피처를 어텐션 맵 생성에 활용 함으로써 콘텐츠 영상의 정보를 손실 없이 결과 영상에 적용한다. StyleFormer [12]는 스타일 뱅크와 어텐션 맵을 결합시켜 다양한 스타일이 적용한 결과 영상을 생성한다. StyTr^2[13] 은 어텐션 개념뿐만 아니라 트랜스포머 프레임워크를 활용하여

영상 내 부분 영역 간의 연관 관계를 학습하여 스타일 패턴 적용에 사용한다.

2. VGG Encoder

VGGNet [13]은 2015년에 이미지 인식을 위해 제안된 CNN 구조로 16개, 19개의 레이어로 이루어진 VGG-16과 VGG-19, 두 개의 모델이 존재한다. 이 모델은 3x3 합성곱 레이어 (Convolution), Relu 활성화 함수, 맥스 풀링, 완전 연결 (Fully Conncted) 레이어로 이루어져 있고 채널 수는 64, 128, 256, 512 순으로 레이어가 증가 됨에 따라 증가한다. VGGNet은 이미지 인식을 위하여 설계된 모델이지만 최근에는 스타일 전이 분야에서는 영상의 특성을 추출하는 목적으로 활용되고 주로 VGG-19 모델이 사용된다.

3. Self Attention

Self attention 기술은 최근 영상 처리와 자연어 처리 분야에서 적극적으로 활용되고 있다. 영상 처리에서는 두 영상 혹은 하나의 영상 내의 포지션간의 상관관계를 구하는 데 사용되고 이 상관관계를 어텐션 맵이라고 한다

스타일 전이 기술에서는 콘텐츠 영상과 스타일 영상 간의 상관관계를 학습하는 과정에서 사용되어 스타일을 적용할 영역이나 관련이 높은 스타일을 선택하는 역할을 한다.

III. Attention based Style Transfer methods

Attention 기술기반으로 스타일 전이 (Style Transfer) 기술 중 대표적인 두 기술은 SANet과 AdaAttN이다. 3장에서는 이 두 기술의 내용과 성능 분석을 소개한다. 콘텐츠 영상은 I_c ,스타일 영상을 I_s, 스타일 변환된 결과 영상을 I_{cs}로 표기한다.

1. SANet (Style-Attentional Network)

SANet[6]은 self attention 기법을 사용해 콘텐츠 영상 내의 피처와 스타일 영상 내의 피처 간의 상관관계를 학습하고 스타일 피처를 콘텐츠 피처에 따라 재배열 한다. Fig.2는 SANet 프레임워크의 구조를 보여준다. SANet 프레임워크는 VGG 인코더, 두 개의 SANet 모듈과, 디코더로 구성되어있다. 콘텐츠 영상, I_c와 스타일 영상 I_s는 VGG 인코더를 통과한 후 피처 맵 F_c와 F_s을 생성한다.

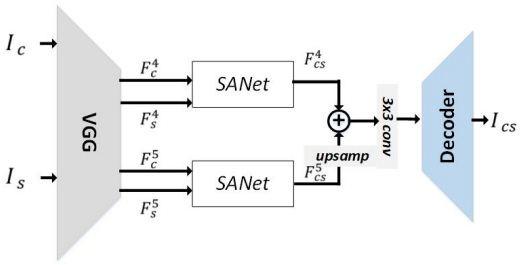

Fig. 2. Framework of SANet

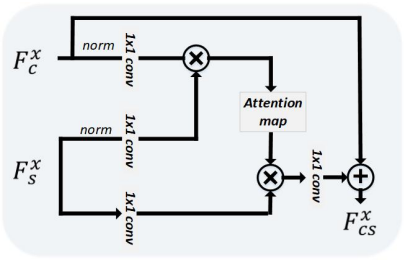

Fig. 3. SANET module

F_c와 F_s는 피처가 추출된 VGG 인코더 레이어에 따라 구분된다. VGG 레이어 중 Relu_a_1에서 추출된 피처맵을 F^a로 정의하고 F_c^4,F_c^5는 Relu_4_1과 Relu_5_1에서 추출된 피처맵을 정의한다. 각각의 SANet 모듈에는 F_c와 F_s가 입력되어 새로운 피처맵 F_{cs}을 생성한다. F_c^4와F_s^4가 입력된 첫 번째 SANet 모듈에서는 F_{cs}^4가 생성되고 두 번째 SANet 모듈에는 F_c^5와 F_c^5가 입력되어 F_{cs}^5가 생성된다. 디코더에 입력되는 최종 피처맵 F_{csc}는 아래 수식에 따라 생성되고 $conv_{3x3}$(3x3 컨벌루션 연산)은 F_{cs}^4와 F_{cs}^4를 병합시키는 역할을 수행한다.

$$F_{\mathrm{csc}} = conv_{3x3}(F_{cs}^4 + upsampling(F_{cs}^5)).$$

$upsampling$ 연산은 F_{cs}^5의 피처맵의 크기를 F_{cs}^4와 맞춰주기 위해 수행된다.

SANet 모듈의 구조는 Fig.3과 같다. F_c와 F_s는 평균 분산 기반 정규화 과정과 1x1 컨벌루션을 거친 후 행렬 곱 연산에 적용되어 어텐션 맵을 생성한다. 획득된 어텐션 맵은 컨텐츠 영상의 피처맵과 스타일 영상의 피처맵 간의 매핑 관계를 학습한다. 획득된 어텐션 맵이 F_s와 곱해지는 과정에서 컨텐츠 특성과 상관관계가 높은 스타일 영상의 피처들이 선택된다. 이 결과는 1x1 컨벌루션을 거친 후 F_c와 결합하여 SANet의 출력 값인 F_{cs}를 생성한다.

학습 과정에서 사용된 손실 함수로는 영상 구조의 비슷한 정도를 측정하는, L_c 와 스타일의 차이를 측정하는 L_s 영상 전체와 부분 영역의 통계적 분포를 동시에 비교하는

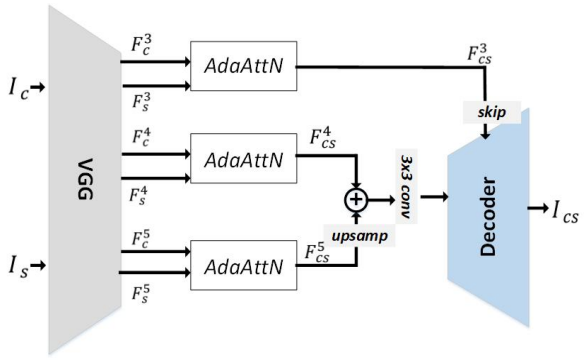

Fig. 4. Framework of AdaAttN

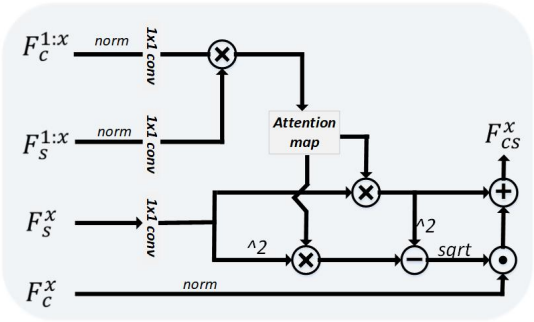

Fig. 5. AdaAttN Module

$L_{identyty}$가 사용되었다.

SANet은 self attention 기법을 이용해 해당 콘텐츠 영상에 최적화된 스타일 피처를 선택하여 스타일 전이 기술의 성능을 향상했고 스타일 영상의 특성을 적용하는데 효과적인 결과를 보여준다. 하지만 Fig.8에서 확인 할 수 있듯이 배경의 색상 분포가 안정적이지 않고 콘텐츠 영상의 디테일을 유지하지 못하는 결과를 보여준다.

3.2. AdaAttN (Adaptive Attention Normalization)

AdaAttN [11] 기술은 SANet 기술을 기반으로 구성되었다. Fig·4는 VGG 인코더와 3개의 AdaAttN 모듈 그리고 디코더로 구성된 AdaAttN의 전체 프레임워크를 보여준다. AdaAttN에서 SANet 기술의 한계점을 극복하고 성능을 향상하기 위하여 변경된 부분들은 다음과 같다.

VGG 인코더의 Relu4_1와 Relu5_1 레이어에서 추출된 피처 맵 을 사용한 SANet과 달리 AdaAttN에서는 Relu3_1에서 추출된 피처 맵, F_c^3,F_s^3도 추가로 사용한다. F_c^3와 F_s^3는 디코더의 중간 레이어의 입력 데이터에 결합(concatenate)되어 스킵 (skip) 연결의 역할을 수행한다. 이러한 스킵 연결은 VGG 인코더 Relu3_1에서 피처 맵을 디코더 중간 레이어에 직접 전달 줌으로써 F_c^3이 포함하고 있는 컨텐츠 영상의 구조적인 디테일을 보존할 수 있게 한다.

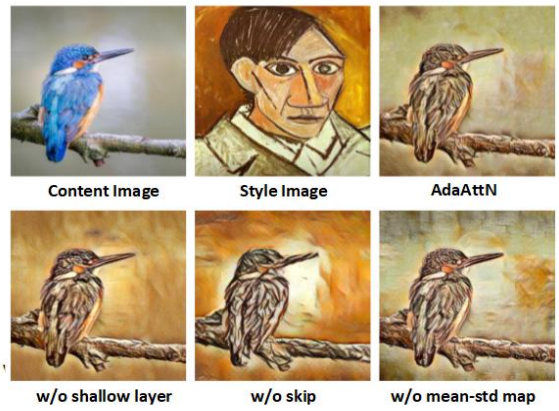

Fig. 6. Analysis of AdaAttN

SANet 모듈과 같이 AdaAttN 모듈도 F_c와 F_s를 입력받아 F_{cs}를 출력한다. 하지만 SANet에서 어텐션 맵 생성에 피처맵 F_*^x를 사용하는 반면 AdaAttN 모듈에서는 $F_*^{1:x}$을 사용한다 (Fig. 5). $F_*^{1:x}$는 아래의 수식과 같이 정의되고 여기서 D_x는 피처 맵을 F_*^x와 같은 크기로 down sampling 하는 연산,$\oplus$ 은 병합(concatenation) 연산을 의미한다.

$$F_*^{1:x} = D_x(F_*^1) \oplus D_x(F_*^2) \oplus \cdots \oplus D_x(F_*^x)$$

이처럼 하위 레이어 피처(shallow feature)까지 어텐션 맵 생성에 사용함으로써 컨텐츠 영상의 디테일까지 고려하여 스타일 피처맵이 생성되고 결과 영상에 컨텐츠 영상의 구조와 디테일이 효과적으로 반영되는 결과를 가져온다.

또한, AdaAttN 모듈은 최종 피처맵을 생성하는 과정에서 평균과 표준편차 맵 (mean-std map)을 (Fig.5) 어텐션 맵을 가중치로 적용하여 결과 영상이 스타일 영상의 색상 분포를 고려하여 생성되도록 한다.

Fig.6은 AdaAttN의 결과 영상과 AdaAttN 제공하는 단위 기능의 효과를 보여준다. AdaAttN의 결과 영상에 비해 스킵(skip) 연결과 하위 레이어 피처 사용이 없는 영상은 디테일의 표현이 부족하거나 색상 분포가 스타일 영상과 차이가 있음을 확인할 수 있다. 평균과 표준편차 맵을 사용하지 않은 영상도 AdaAttN의 결과 영상과 비교해 배경의 색상 분포가 안정되지 않음을 확인 할 수 있다.

사용된 손실 함수로는 콘텐츠 영상과의 구조적 차이를 측정하는 L_{cn}, 전체 영상의 스타일 차이를 하는 L_{gs}와 부분 영역의 스타일 차이를 측정하는 L_{lf}이 사용되었다.

이처럼 AdaAttN은 SANet 프레임워크에서 몇 가지 변화를 주어 콘텐츠 영상의 디테일을 결과 영상에 효과적으로 반영한다. 하지만 Fig.8에서 볼 수 있듯이 스타일 표현

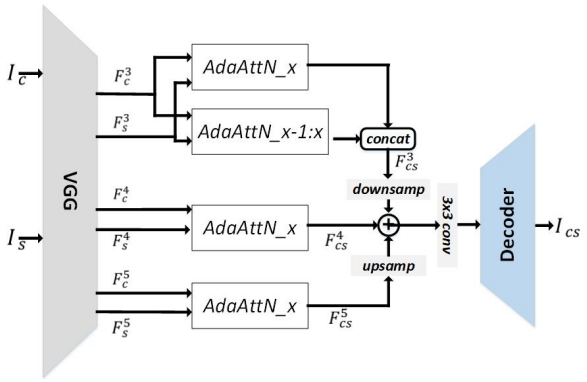

Fig. 7. Framework of Proposed Methods

이 제한적이라는 한계도 존재한다. 두 번째 스타일 이미지의 경우 원형 패턴을 특징으로 가지고 있지만 AdaAttN의 결과 영상에는 원형 패턴을 찾을 수가 없다.

IV. Proposed Methods

위와 같이 SANet과 AdaAttN은 어텐션 기술을 효과적으로 적용하여 스타일 전이 기술의 성능 향상을 이루었다. 하지만 SANet의 경우 콘텐츠 영상의 구조를 보존하지 못하고 영상 전체의 색상 분포가 안정되지 못한 문제점이 존재한다. AdaAttN은 콘텐츠 영상의 정보의 보존에 치우쳐 스타일을 충분히 결과 영상에 반영하지 못한 한계가 있다. 이 두 기술의 문제점을 해결하여 콘텐츠 영상의 구조 보존과 스타일 영상의 패턴 적용을 동시에 고려하고자 다음의 방법을 제안한다.

제안하는 기술은 Fig. 7에서와같이 AdaAttN의 프레임워크를 기반으로 하고 있고 변형된 AdaAttN 모듈인 AdaAttN_x와 AdaAttN_x-1:x를 사용한다.

AdaAttN_x 모듈은 AdaAttN 모듈의 $F_c^{1:x}$와 $F_s^{1:x}$을 F_c^x와 F_s^x변경한 것이다. 하위 레이어의 피처 맵까지 어텐션 맵 생성에 사용하는 AdaAttN와 달리 해당 레이어의 피처 맵만을 사용하여 콘텐츠 영상의 디테일을 지나치게 보존하여 스타일을 반영하지 못하는 상황을 방지한다. AdaAttN_x-1 모듈은 AdaAttN 모듈의 $F_c^{1:x}$와 $F_s^{1:x}$을 $F_c^{x-1:x}$와 $F_s^{x-1:x}$로 변경한 것이다.

AdaAttN에선 F_c^3와 F_s^3이 디코더 중간 레이어에 직접 입력되지만 제안하는 기술에서는 F_*^3이 F_*^4, F_*^5와 더해져 디코더에 입력된다. 이는 하위 레이어 정보가 직접 입력되었을 때 콘텐츠 영상의 디테일이 지나치게 보존되는 것을 방지하면서도 하위 레이어 정보를 다른 피처맵 정보와 함께 디코더에 전달하여 콘텐츠 영상의 구조 정보를 전달하려는 방법이다. 채널이 256개인 F_*^3을 채널이 512개인 F_*^4, F_*^5와 맞춰주기 위해서 F_*^3 을 AdaAttN_x과 AdaAttN_x-1:x에 각각 입력하여, 두 개의 피처 맵, F_{cs1}^3 F_{cs2}^3 을 생성한다. F_{cs1}^3 과F_{cs2}^3 은 결합(concatenate) 되어 채널이 512인 F_{cs}^3 을 생성한다. F_*^4와 F_*^5가 각각 AdaAttN_x에 입력되어 생성된 F_{cs}^4, F_{cs}^5 는 아래 식과 같은 방법으로 F_{cs}^3 와 더해진 후 3x3 컨벌루션 연산을 거쳐 디코더에 입력된다. 아래 식에서 D_4와 U_4는 F_{cs}^3 와 F_{cs}^5 를 F_{cs}^4 의 크기와 맞추기 위한 down sampling과 up sampling을 의미한다. w3, w4, w5는 각 피처 맵의 가중치를 뜻한다.

$$F_{\mathrm{csc}_m} = (w3 * D_4(F_{cs}^3) + w4 * F_{cs}^4 + w5 * U_4(F_{cs}^5))$$

$$F_{\mathrm{csc}} = conv_{3x3}(F_{\mathrm{csc}_m})$$

V. Experiments

제안하는 모델의 훈련을 위하여 MS-COCO [14] 데이터 세트를 콘텐츠 영상의 훈련 데이터로 와 WikiArt [15] 데이터 세트를 스타일 영상의 훈련 데이터로 사용하였다. MS-COCO 데이터 세트와 WikiArt 데이터 세트는 각각 약 80,000개의 영상으로 이루어져 있다. 훈련 시 옵티마이저는 Adam을 사용하였고 학습률은 0.0002, 배치 사이즈는 8을 적용하였다. 훈련과 테스트 과정은 Nvidia Tesla V100 GPU를 사용하여 수행되었다. 제안된 기술은 파이토치로 구현된 AdaAttN 코드를 기반으로 구현되었다. IV 장의 마지막 부분에 설명된 $F_{cs}^3, F_{cs}^4, F_{cs}^5$ 가 더해지는 과정에서 사용된 가중치 w3, w4, w5는 각각 0.1, 0.3, 0.6으로 설정되었다. 가중치는 하위 레이어 보다 상위 레이어의 가중치를 크게 하도록 설정되었고 이는 상위 레이어가 스타일 표현과 연관성이 높으므로 스타일 보존을 우선으로 하기 위함이다.

Fig·8은 SANet과 AdaAttN, 그리고 본 논문에서 제안하는 기술의 결과 영상을 보여준다. 검은색 박스로 표시된 영역은 흰색 박스로 표시된 영역을 확대한 것이다. Fig. 8.의 SANet 결과 영상을 보면 두 개의 결과 영상 모두 배경의 색상 분포가 콘텐츠 영상과 차이가 크고 안정적이지 못함을 확인할 수 있다. 특히 콘텐츠 영상의 눈 부분과 건물 일부 디테일의 변형이 심하고 생략되는 부분 (건물의 창문 등

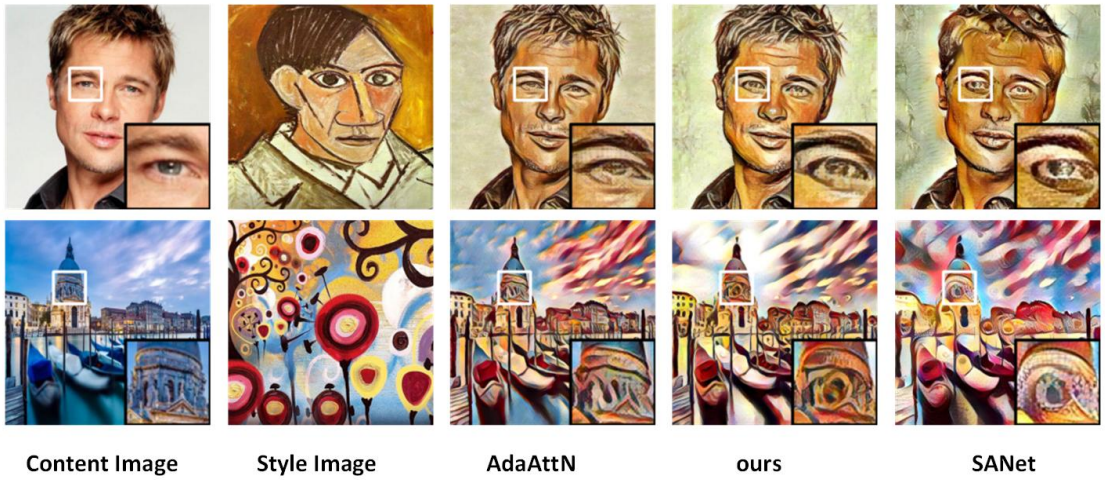

Fig. 8. Comparison of SANet, AdaAttN, proposed methods

) 이 존재함을 확인할 수 있다. AdaAttN는 콘텐츠 영상의 정보를 효과적으로 결과 영상에 전달하지만, 첫 번째 결과 영상의 경우 스타일 영상의 특성인 굵은 선(스트로크)과 명암 대비가 충분히 반영되지 못했고 두 번째 테스트의 경우는 스타일 영상의 특징인 원형의 패턴이 반영되지 않았음을 확인 할 수 있다. 제안한 기술의 결과 영상의 경우 콘텐츠 영상의 구조를 보존하면서도 (눈의 모양이나 건물의 디테일) 스타일 영상의 특징 (스트로크의 특성이나 원형 패턴의 표현 등)을 반영하고 있어 SANet과 AdaAttN 기술의 각각의 장점을 취하며 균형을 맞추고 있음을 보여준다.

Fig.9는 SANet과 AdaAttN, Avatar-Net. StyleFormer 등의 기존 기술과 본 논문에서 제안하는 기술의 결과 영상을 보여준다, Avatar-Net의 경우 콘텐츠 영상의 구조 반영과 스타일 영상의 특징 반영의 두 가지 기준을 만족시키지 못함을 확인할 수 있다. StyleFormer의 경우는 스타일을 잘 표현하고 있지만 콘텐츠 영상의 구조 반영이 미비한 영상 결과가 존재한다. 특히 세 번째 결과 영상에서는 얼굴의 디테일이 거의 반영 되지 못했고 여섯 번째 결과 영상에서는 새의 모양을 표현하지 못하고 있다. 제안된 기술은 대부분의 테스트 영상에서 콘텐츠 영상의 구조와 디테일에 스타일이 효과적으로 적용되어 안정적인 성능을 보여준다.

VI. Conclusions

본 논문에서는 스타일 전이 기술의 사용성 향상을 위하여 최근 활발히 연구되는 어텐션 기반 기술들은 분석하고 각 기술의 단점을 극복하는 새로운 모델을 제안하였다. 제안하는 기술은 스타일 전이 기술의 성능에서 가장 중요한 기준인 콘텐츠 영상의 정보 유지와 스타일 영상의 특징 적용 간의 안정적인 균형을 찾는 결과 영상을 생성한다. 본 연구의 새로운 연구 방향으로는 동영상에 적용하기 위한 프레임 간의 연결성을 고려한 스타일 전이 기술의 연구가 고려된다.

ACKNOWLEDGEMENT

This work was supported by INHA TECHNICAL COLLEGE Research Grant in 2022

REFERENCES

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414-2423, 2016. DOI: 10.1109/CVPR.2016.265

[2] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S Lempitsky, “Texture networks: Feed-forward synthesis of textures and stylized images,” Proceedings of International Conference on Machine Learning, pp.1349-1357, 2016. DOI:10.48550/arXiv.1603.03417

[3] K. Simonyan and A. Zisserman.,“Very deep convolutional networks for large-scale image recognition,” Proceedings of International Conference on Learning Representation, 2015. DOI:10.48550/arXiv.1409.1556

[4] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.Yang, “Universal style transfer via feature.transforms,” Proceedings of Neural Information Processing Systems, pp.386-396, 2017. DOI: 10.48550/arXiv.1705.0808

[5] X. Huang and S. Belongie, “Arbitrary style transfer in real-time

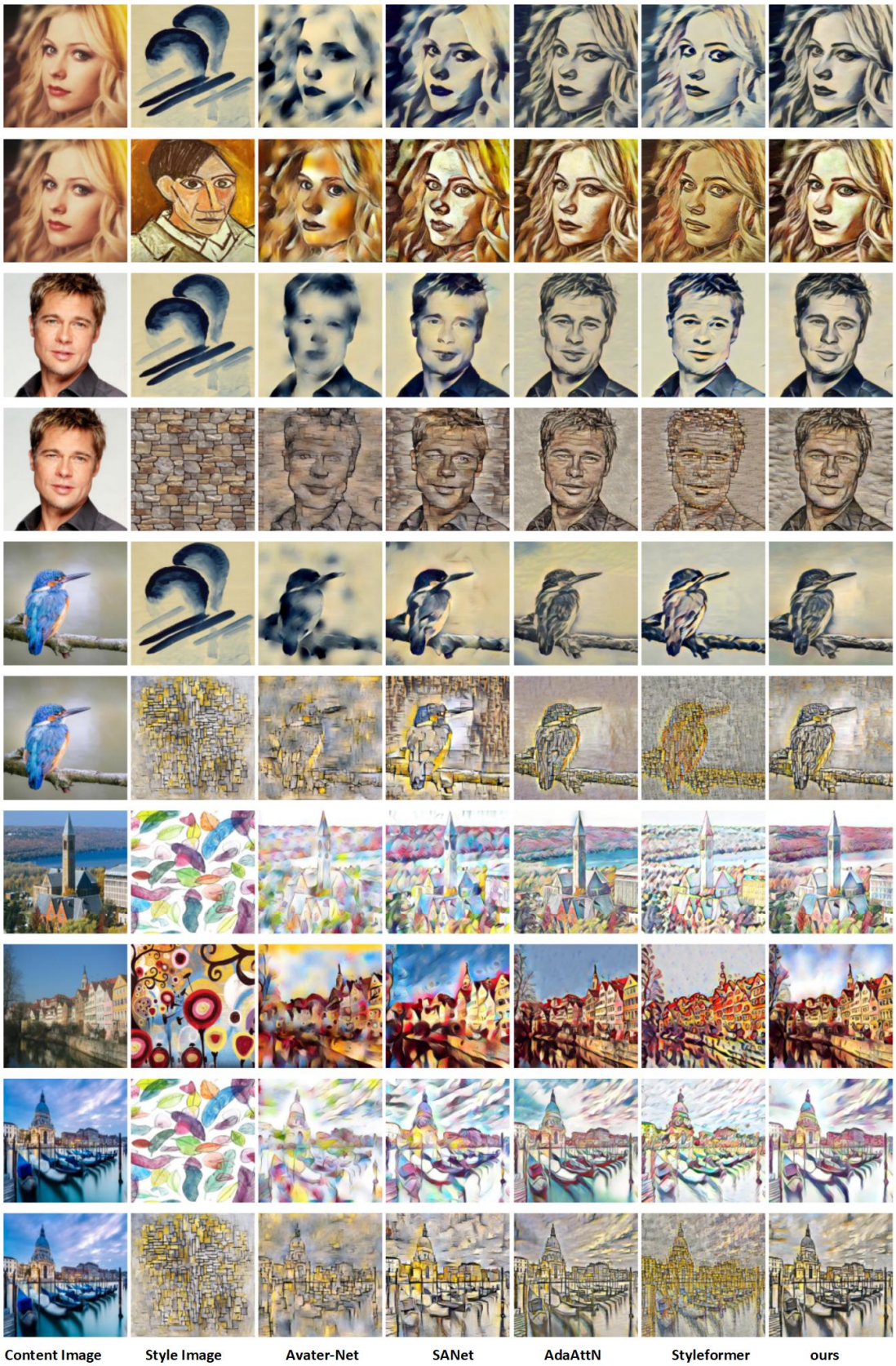

Fig. 9. Comparison of result images

with adaptive instance normalization," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1501-1510, 2017. DOI: 10.1109/ICCV.2017.167

[6] D. Y. Park and K. H. Lee, "Arbitrary style transfer with style-attentional networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.5880-5888, 2019. DOI:10.1109/CVPR.2019.00603

[7] L. Sheng, Z. Lin, J. Shao, and X. Wang, "Avatar-net: Multi-scale zero-shot style transfer by feature decoration," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8242-8250, 2018. DOI: 10.1109/CVPR.2018.00860

[8] Y. Yao, Ji. Ren, X. Xie, W.Liu, Y. Liu, and J.Wang, "Attention-aware multi-stroke style transfer," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1467-1475, 2019. DOI: 10.1109/CVPR.2019.00156

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N Gomez, Lu. Kaiser, and I. Polosukhin, "Attention is all you need," Proceedings of Neural Information Processing Systems, pp.5998-6008, 2017. DOI:0.48550/arXiv.1706.03762

[10] Y. Deng, F. Tang, W. Dong, W. Sun, F. Huang, and C. Xu, "Arbitrary style transfer via multi-adaptation network," Proceedings of ACM Multimedia, pp.2719-2727, 2020. DOI:10.1145/3394171.3414015

[11] S. Liu, T. Lin, D. He, F. Li, M. Wang, X. Li, Z. Sun, Q. Li, and E. Ding, "AdaAttn: Revisit attention mechanism in arbitrary neural style transfer," Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021. DOI: 10.1109/ICCV48922.2021.00658

[12] X. Wu, Z. Hu, L. Sheng, and D. Xu, "Styleformer: Real-time arbitrary style transfer via parametric style composition," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.14618-14627, 2021. DOI: 10.1109/ICCV48922.2021.01435

[13] Yingying Deng, Fan Tang, Weiming Dong Member, IEEE, Chongyang Ma, Xingjia Pan, Lei Wang, Changsheng Xu, "StyTr2 :Image Style Transfer with Transformers," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2022. DOI: 10.1109/CVPR52688.2022.01104

[14] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L.Zitnick, "Microsoft coco: Common objects in context," Proceedings of European Conference on Conputer Vision, pp. 740-755. 2014. DOI: 10.1007/978-3-319-10602-1_48

[15] F. Phillips and B. Mackintosh, "Wiki art gallery, inc.: A case for critical thinking." Accounting Education, Vol. 26, no.3, pp.593-608, 2011. DOI: 10.2308/iace-50038

Authors

Junghye Min received the B.S. degree in mathematics from Ewha Women's University, Seoul, Korea, in 1995, and the M.E. and Ph.D. degrees in computer science and engineering from Pennsylvania State University, U.S.A., in 2003 and 2005, respectively. From 2005 to 2021 she was a Principal Engineer with Samsung Research, Samsung Electronics, Seoul, Korea. Dr. Min joined the faculty of the Department of Computer Science at Inha Technical College, Inchoen, Korea in 2022 and is currently an assistant Professor. Her research interests include image enhancement, image style transfer, and pattern recognition.