

웨이블릿 혼합 모듈: 웨이블릿 변환을 이용한 네트워크 내 고주파 성분 보존

김민우, *조남익

서울대학교 전기정보공학부, 뉴미디어통신공동연구소
mwk0614@ispl.snu.ac.kr, *nico@snu.ac.krWavelet Mix Module: Preserving High-Frequency in Network
using Wavelet Transform

Min Woo Kim, *Nam Ik Cho

Department of ECE, INMC, Seoul National University

요 약

본 논문에서는 ‘스케치로부터 RGB 이미지로의 변환’을 수행하는 웨이블릿 기반의 네트워크에서 생성된 이미지 품질을 높이기 위해, 네트워크가 저주파수에 편향되어 학습이 되는 것을 완화하고자 Wavelet Mix Module(WMM)을 제안하였다. WMM 은 UNet 구조의 skip-connection 과정에 적용되며, 웨이블릿 변환을 사용하여 인코더 특성으로부터 세부값을 추출하여 디코더 특성으로 전달함으로써 네트워크 내에서 고주파 성분이 보존되도록 한다. WMM 이 적용된 네트워크로부터 생성된 이미지는 정량적 및 정성적인 결과가 개선됨을 실험을 통해 확인하였다.

1. 서론

스케치는 추상적이지만 간결하고 직관적인 표현이 가능하기 때문에, 사람들이 어떤 대상을 그림으로 묘사할 때 가장 많이 사용하는 기법 중 하나이다. 따라서 스케치와 관련된 컴퓨터 비전 연구가 많이 이루어지고 있다. 대표적으로 ‘스케치 클래스 분류, 스케치 기반 물체 검색, 스케치로부터 RGB 이미지로의 변환’ 등이 있으며, 본 논문에서는 ‘스케치로부터 RGB 이미지 변환’에 관하여 연구를 진행하였다. ‘스케치로부터 RGB 이미지로의 변환’은 스케치 데이터 분포를 RGB 이미지(이하 이미지) 데이터 분포로 매핑하는 과정이며, 생성 모델이 주로 사용된다. 하지만 합성곱 신경망으로 구성된 생성 모델은 저주파수를 위주로 학습한다는 것이 알려져 있다 [1]. 즉, 생성 모델은 고주파수를 학습함에 어려움이 있고, 이로 인해 발생하는 주파수 차이는 이미지 품질 저하를 야기한다.

위 주파수 차이를 극복하기 위해서, 네트워크 관점에서 다양한 아이디어가 제시되었다. 예를 들면, [2]에서는 합성곱

신경망과 multi-spectral channel attention(MMCA)을 사용하여, skip-connection 을 통해 고주파수를 보다 잘 전달하고자 하였다. 또한 [3]에서는 학습된 마스크를 이용하여 이미지를 저주파수, 중간 주파수와 고주파수로 분리한 후, 모델의 학습 역량을 고주파수에 큰 비중으로 할당하여 성능과 속도를 향상시켰다.

위에서 제안된 방법들은 학습 가능한 변수와 네트워크를 통해 고주파수를 다루고자 하였지만, 학습 설정과 환경에 따라서 학습 내용이 달라질 수 있고 항상 고주파수를 주안점으로 학습을 하는지 보장할 수 없다. 따라서 본 논문에서는 웨이블릿 변환을 이용하여 직접적으로 고주파수를 추출하여 네트워크 내에서 전달하는 모듈인 Wavelet Mix Module(WMM)을 제안한다. 이 모듈은 Unet[4]의 skip-connection 과정에 적용되어 인코더 특성으로부터 추출한 고주파 성분을 디코더 특성으로 전달함으로써, 네트워크가 저주파수에 편향되어 학습되는 것을 완화하도록 설계되었다. 이후 실험을 통해, 제안된 모듈이 네트워크에 적용되면 정량적 및 정성적 결과가 모두 향상되는 것을 확인하였다.

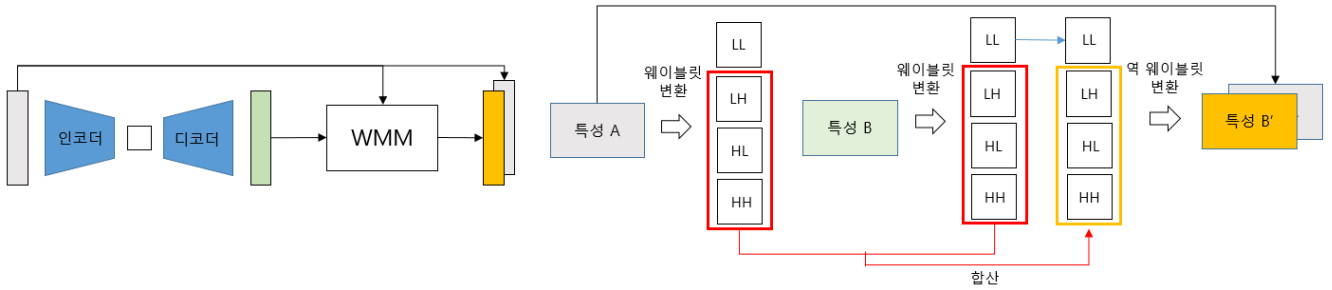


그림 1. 제안된 Wavelet Mix Module(WMM)이 네트워크 내에서 위치하는 부분(좌)과 WMM 에서 인코더 특성의 고주파 성분이 디코더 특성으로 전달되는 과정(우)을 나타낸다. 웨이블릿 변환을 거치면 저주파 정보를 담고 있는 근사값과 고주파 정보를 담고 있는 세부값이 나오는데, 근사값은 LL 로, 세부값은 LH, HL, HH로 표기하였다.

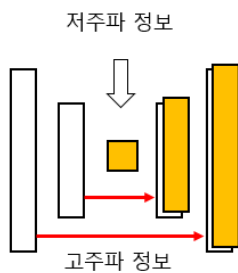


그림 2. 네트워크 구조 개요.

2. Wavelet Mix Module

기존 논문[5]에서도 웨이블릿 변환을 사용하여 네트워크를 설계하였지만, 고주파에 초점을 맞추어 설계되지는 않았다. 따라서 본 절에서는 네트워크 내에서 고주파 성분이 보존되도록 하는 네트워크 모듈을 제안한다. 설계에 필요한 핵심 가정은, 네트워크에서 가장 낮은 차원(높이와 너비 기준)을 가지는 특성이 저주파 정보를 충분히 가지고 있다는 것이다(그림 2, 중간 화살표). 따라서 skip-connection 과정에서는 인코더 특성으로부터 고주파 성분을 추출하여 디코더 특성으로 전달한다(그림 2, 붉은색 화살표).

기존 Unet 구조는 인코더 특성이 그 자체로 디코더 특성과 연결(concatenation)되어 다음 합성곱 신경망으로 넘어가지만, WMM 은 인코더 특성과 디코더 특성을 모두 입력으로 받아 변형된 디코더 특성을 생성하여 이를 인코더 특성과 연결한다(그림 1, 좌). 이 과정은 그림 1의 우측에 묘사되어 있다. 먼저 WMM 에 입력으로 들어간 인코더 특성(특성 A)과 디코더 특성(특성 B)에 웨이블릿 변환을 적용하여 저주파 정보를 담고 있는 근사값과 고주파 정보를 포함하는 세부값으로 분리한다. 그 후 특성 A 에서 추출된 세부값을 특성 B 의 세부값에 더하여,

특성 A 의 고주파 정보를 반영하는 변형된 세부값을 생성한다. 그리고 이 세부값과 특성 B 의 근사값에 역 웨이블릿 변환을 수행하여 특성 B'을 만든다. 그리고 특성 B'과 특성 A 를 연결하여 다음 합성곱 신경망으로 전달한다.

3. 실험 결과 및 분석

3.1. 실험 설정

본 논문에서 제안한 모듈을 비교 실험하기 위한 네트워크로, Pix2Pix[6]와 Focal Frequency Loss[7]를 결합한 모델을 선정하였다. Pix2Pix 의 생성자는 Unet 구조로 되어 있는데, skip-connection 과정에 WMM 을 적용하여 실험을 진행하였다. WMM 은 웨이블릿 변환을 포함하고 있으므로 두 개의 초매개변수(웨이블릿 기저, 웨이블릿 분해 수준)가 있는데, 실험에서는 각각을 ‘Haar 웨이블릿 기저’ 및 ‘2’로 설정하였다.

본 논문에서 제안한 네트워크 구조를 실험하기 위한 데이터셋으로, 스케치와 RGB 이미지가 쌍으로 구성된 ‘edges2shoes’[8]를 채택하였다. 이 데이터셋은 스케치를 RGB 이미지로 변환하는 것(그림 3)을 목표로 하며, 학습 샘플 49,825 장과 테스트 샘플 200 장으로 구성되어 있다.

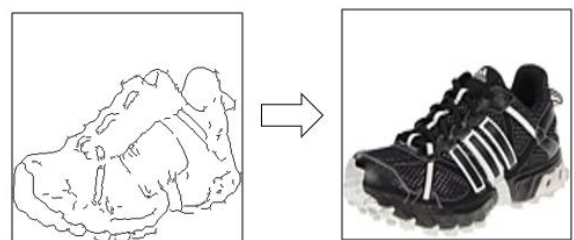


그림 3. ‘edges2shoes’ 데이터셋 구성 예시



그림 4. 정성적 결과 비교. 1 열에서부터 순차적으로, 스케치, 기존 네트워크로부터 생성된 이미지, WMM 이 추가된 네트워크로부터 생성된 이미지와 실제 이미지를 전시하였다. 붉은색 박스로 표시한 부분은, WMM 적용 후 이미지 상에서 특히 개선된 곳을 나타낸다.

3.2. 정량적 결과 비교

실험 결과의 정량적 수치를 측정하기 위해 Frechet Inception Distance(FID)[9]를 사용하였다. FID 는 생성된 이미지 집합과 실제 이미지 집합으로부터

$$FID = \|\mu_X - \mu_Y\|^2 - Tr(\Sigma_X + \Sigma_Y - 2\Sigma_X\Sigma_Y) \dots \text{식}(1)$$

과 같이 계산된다. 식(1)에서 아래첨자 X 는 생성된 이미지 집합, 아래첨자 Y 는 실제 이미지 집합을 의미하고, 기호 μ 는 평균, 그리고 기호 Σ 는 공분산 행렬을 뜻한다. FID 값이 낮을수록 이미지 집합 사이의 통계량(평균, 공분산) 차이가 작다는 것을 의미하기 때문에 생성된 이미지 분포가 실제 이미지 분포와 가깝다는 것이므로 생성 모델의 성능이 좋음을 의미한다. 표 1 은 비교 대상이 되는 기존 네트워크와 WMM 이 적용된 네트워크로부터 생성된 이미지 집합을 토대로 계산된 FID 를 기입한 것이며, WMM 을 적용했을 때 정량적 수치가 개선됨을 확인할 수 있다.

표 1. WMM 여부에 따른 FID. WMM 을 적용하지 않은 기존 네트워크에 대한 결과는 [7]에서 보고된 수치를 인용하였다.

기존	74.359
+ WMM	66.453

3.3. 정성적 결과 비교

기존 네트워크에 WMM 을 적용시켰을 때, 정성적 결과에도 개선이 있는지 확인하였다(그림 4). 그 결과, WMM 이 사용된 네트워크로부터 생성된 이미지에서 이미지 품질을 크게 저하시키는 아티팩트가 존재하는 비율이 감소하였다. 예를 들면, 그림 4 의 각 행에 붉은색 박스로 표시한 부분에 눈에 띄게 큰 아티팩트가 존재했는데, WMM 을 사용하면 부자연스러운 무늬가 완화되는 것을 관찰할 수 있다.

4. 결론

본 논문에서는 웨이블릿 변환을 적용하여 고주파 성분을 직접적으로 네트워크 내에서 보존함으로써 생성된 이미지의 품질을 높이고자, Wavelet Mix Module(WMM)을 제안하였다. 이를 적용하면 생성된 이미지 분포와 실제 이미지 분포가 통계적으로 가까워지고, 품질에 크게 영향을 주는 아티팩트가 줄어드는 등 정량적 및 정성적 결과가 개선됨을 확인하였다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행되었으며, 2022 년도 BK21 FOUR 정보기술

미래인재 교육연구단에 의하여 지원되었음 (2021R1A2C2007220).

참고 문헌

- [1] Nasim Rahaman et al, "On the Spectral Bias of Neural Networks", In International Conference on Machine Learning(ICML), 2019
- [2] SA Magid et al, "Dynamic High-Pass Filtering and Multi-Spectral Attention for Image Super-Resolution", In International Conference on Computer Vision(ICCV), 2021
- [3] Xie Wenbin et al, "Learning Frequency-aware Dynamic Network for Efficient Super-Resolution", In International Conference on Computer Vision(ICCV), 2021
- [4] Olaf Ronneberger et al, "U-Net: Convolutional Networks for Biomedical Image Segmentation", In Medical Image Computing and Computer Assisted Intervention(MICCAI), 2015
- [5] Pengju Liu et al, "Multi-Level Wavelet Convolutional Neural Networks". In Institute of Electrical and Electronics Engineers(IEEE) Access, 2019
- [6] Phillip Isola et al, "Image-to-Image Translation with Conditional Adversarial Networks", In Conference on Computer Vision and Pattern Recognition(CVPR), 2017
- [7] Liming Jiang et al, "Focal Frequency Loss for Image Reconstruction and Synthesis", In International Conference on Computer Vision(ICCV), 2021
- [8] Aron Yu et al, "Fine-grained visual comparisons with local learning", In Conference on Computer Vision and Pattern Recognition(CVPR), 2014
- [9] Martin Heusel et al, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", In Neural Information Processing Systems(NIPS), 2017