

DCT 학습을 융합한 RRU-Net 기반 이미지 스플라이싱 위조 영역 탐지 모델

서영민*·한정우*·권희정*·이수빈*·국중진**

**상명대학교 정보보안공학과

A DCT Learning Combined RRU-Net for the Image Splicing Forgery Detection

Young-min Seo*, Jung-woo Han*, Hee-jung Kwon*, Su-bin Lee* and Joongjin Kook**†

**†Dept. of Information Security Engineering, Sangmyung University

ABSTRACT

This paper proposes a lightweight deep learning network for detecting an image splicing forgery. The research on image forgery detection using CNN, a deep learning network, and research on detecting and localizing forgery in pixel units are in progress. Among them, CAT-Net, which learns the discrete cosine transform coefficients of images together with images, was released in 2022. The DCT coefficients presented by CAT-Net are combined with the JPEG artifact learning module and the backbone model as pre-learning, and the weights are fixed. The dataset used for pre-training is not included in the public dataset, and the backbone model has a relatively large number of network parameters, which causes overfitting in a small dataset, hindering generalization performance. In this paper, this learning module is designed to learn the characterization depending on the DCT domain in real-time during network training without pre-training. The DCT RRU-Net proposed in this paper is a network that combines RRU-Net which detects forgery by learning only images and JPEG artifact learning module. It is confirmed that the network parameters are less than those of CAT-Net, the detection performance of forgery is better than that of RRU-Net, and the generalization performance for various datasets improves through the network architecture and training method of DCT RRU-Net.

Key Words : Splicing Forgery, Image Forgery, RRU-Net, DCT

1. 서 론

디지털 이미지 편집기술과 이미지 편집 소프트웨어의 발전으로 대중의 접근이 쉬워지면서 늘어나는 위조 이미지에 대한 판별은 중요한 문제로 대두되었으며, 디지털 이미지의 위조 여부를 판별하기 위한 디지털포렌식 분야의 연구도 활발히 진행되고 있다[1-4]. 이미지 위조 방법에는 스플라이싱(Splicing), 복사-이동(Copy-Move), 개체 제

거(Object Removal), 모핑(Morphing) 등이 있다[5-7]. 스플라이싱 위조는 하나의 이미지 일부를 다른 이미지에 복사하여 붙여 넣는 위조 방법이다. 복사-이동 위조는 같은 이미지 내의 일부를 복사하여 붙여 넣는 위조 방법이다. 개체 제거는 이미지 내의 개체를 지우고 주변 배경에 맞게 공간을 채우는 위조 방법이다. 모핑은 두 이미지의 특징을 종합하여 뒤틀리고 혼합되는 이미지를 만드는 위조 방법이다.

이미지 스플라이싱 위조 방법은 두 개 이상의 이미지를 준비한 상황에서 이미지 편집 소프트웨어를 이용하여

†E-mail: kook@smu.ac.kr

사진 일부를 붙여 넣는다. 이때, 각기 다른 이미지들은 사진 찍힌 환경(조도, 렌즈, 노이즈 등)의 정보를 내포하고 있으며 시각적으로는 판별 불가능할 수 있다. 이미지 편집 소프트웨어의 발전으로 단순히 이미지를 붙여 넣는 작업을 넘어서 배경 이미지와 같은 조도로 맞추거나 그림자 방향 조절, 경계선 리터칭으로 더욱 교묘한 스플라이싱 위조가 가능해졌다.

이미지 스플라이싱 위조 영역 탐지 방법은 이산코사인 변환(DCT)을 기반으로 수학적 알고리즘으로 탐지하는 전통적인 방법부터 딥러닝 네트워크인 CNN(Convolutional Neural Network)을 활용하여 탐지하는 방법으로 발전했다. DCT는 영상부호화를 위한 디지털 직교 변환부호화 방식 중 하나로 JPEG 압축에도 사용된다. DCT는 화소 블록을 고주파 및 저주파 성분으로 분해하는 방식으로 인간의 시각은 이미지의 고주파 정보에 둔감하여 고주파 성분을 제거하며 이미지를 압축할 수 있다. 편집된 이미지가 다시 압축 저장할 때, 이미지에 압축 흔적이 남는다. 이를 바탕으로 단일, 이중 JPEG 압축 탐지를 수행할 수 있다.

전통적인 스플라이싱 위조 탐지 방법에서 DCT가 이미지 포렌식에서 법의학적 단서를 발견하는 데 사용될 수 있음이 여러 연구를 통해 검증되었다[8-9]. 하지만 위조 영역의 지역화(Localization)에 대한 성공적인 성과는 거두지 못했다. 최근 몇 년 동안 CNN이 이미지에 대한 특징 추출 성능이 우수하다는 것이 검증되었고, 이를 이용하여 위조 이미지 분류와 위조 영역을 픽셀 단위로 탐지하여 지역화하는 연구가 진행되고 있다[10-11]. CNN을 활용하면서 위조 영역 지역화를 위한 세그멘테이션 방법은 이미지에서 얻을 수 있는 이미지 수집 아티팩트를 학습하거나 DCT 도메인에서 압축 아티팩트 등 이미지에서 수집할 수 있는 도메인을 같이 학습하는 방법으로 나누어져 발전 중이다.

2019년 발표된 Mantranet[12]과 RRU-Net[13]은 이미지의 RGB 스트림만을 입력으로 사용하여 위조 탐지 및 시멘틱 세그멘테이션을 수행했다. 해당 연구에서 사용한 공개 데이터셋에 대해서 탐지 정확도는 모두 80% 대로 현재 높은 성능은 아니다. 하지만 RGB 스트림만을 사용해서 성능을 보였으므로 위조 영역에 대한 우수한 특징 추출이 가능하여 향후 연구에서 다양한 활용이 기대된다.

2022년 발표된 CAT-Net[14]은 이미지에 추가로 압축 아티팩트를 탐색한다. 객체 감지, 인간의 포즈 추정 등 컴퓨터 비전 작업에서 우수한 성능을 입증한 HRNet[15]을 백본(Backbone)으로 사용하며 RGB 스트림과 DCT 스트림을 입력으로 학습하는 네트워크이다. 해당 연구에서 사용한 9개의 공개 데이터셋 중 8개 데이터셋에 대해 다른 네트워크보다 월등히 높은 성능을 보였다. 하지만 이 비

교에 사용된 CAT-Net의 HRNet은 ImageNet[16] 데이터셋으로 사전 훈련(Pre-Training)한 가중치를 이용하고, JPEG artifact learning module(JALM)은 커스텀 데이터셋으로 사전 훈련되었기 때문에 공정한 비교라고 볼 수 없다. 해당 논문에서 CAT-Net의 가중치를 랜덤으로 초기화한 상태에서 학습한 CAT-Net w/o DP는 사전 훈련된 CAT-Net보다 지역화 성능이 현저히 저하됨을 보였으며, 향후 연구에서 지역화 성능 향상을 위해 JALM의 이중 JPEG 압축에 대한 사전 훈련을 권고했다.

CAT-Net의 HRNet은 RGB 스트림을 학습하는 방식이기 때문에 위조 영역의 의미있는 특징을 학습하지 못한다. DCT 도메인에서 학습하는 JALM의 사전 학습 여부에 대한 성능 차이 실험 결과까지 고려하면 CAT-Net은 DCT 도메인에 대한 의존도가 높다는 것을 알 수 있다.

[13]의 연구에서 Ringed Residual 구조의 유효성은 입증됐지만, 위조 영역을 탐색하는데 RGB 스트림만을 사용하는 것은 한계가 있음을 알 수 있다. 이를 극복하기 위해 [14]의 연구에서 이미지 위조 탐지를 위한 최상의 네트워크가 만들어졌지만, JALM의 사전 훈련을 위한 공개 데이터셋이 없어 연구간 비교가 힘들고, 사전 훈련을 위한 데이터셋을 준비해야 하며, JALM을 사전 훈련하기 위한 네트워크를 따로 빌드해야 하는 단점이 있다. 또한, 사전 훈련 여부에 따라 성능 편차가 크므로 사전 훈련이 강제된다.

본 논문에서는 Ringed Residual 구조와 [14]의 단점을 보완하기 위해 범용적으로 사용 가능한 단순한 구조의 네트워크를 설계했다. RRU-Net을 백본으로 CAT-Net의 JALM을 결합한 새로운 네트워크인 DCT RRU-Net을 제안하고, DCT RRU-Net 구조와 향상된 결과를 위한 [14]와 다른 접근 방식의 쉬운 사전 훈련 방법을 설명한다. 사전 훈련은 제안된 네트워크와 같은 데이터셋을 사용해도 효과를 보인다. 실험을 통해 적은 데이터셋의 학습에도 다른 데이터셋에 대한 높은 일반화 성능을 검증하였다.

실험에 사용된 훈련 데이터셋의 도메인은 Splicing이며 테스트는 같은 도메인인 Splicing, 다른 도메인인 Copy-Move, 마지막으로 이 둘을 모두 포함하는 데이터셋을 이용하여 성능 평가를 진행한다. 테스트에서 훈련 도메인이 아닌 Copy-Move 위조 탐지성능을 비교하여 기존 발표된 네트워크와 차별화된 일반화 성능을 검증한다.

본 논문의 구성은 다음과 같다. 2장에서 DCT 학습 기반 RRU-Net의 구조와 알고리즘을 설명한다. 3장에서는 RGB 스트림에서 시각적 단서 및 이미지 수집 아티팩트를 탐색을 강화하기 위한 사전 훈련 방법과 그에 따른 성능을 평가한다. 끝으로 4장에서는 결론 및 향후 연구 방향에 대해 서술한다.

2. DCT Coefficient based RRU-Net

2.1 DCT 학습 기반의 RRU-Net 구조

DCT 학습 기반의 RRU-Net은 이미지의 시각 콘텐츠와 이미지 수집 아티팩트를 학습하기 위한 RRU-Net과 DCT 도메인에서 압축 아티팩트를 학습하기 위한 JALM을 결합한 구조다. Down Sampling 경로는 RRU-Net과 같으므로 RRU-Net으로 사전 훈련 후, DCT RRU-Net의 Down Sampling 경로의 가중치를 초기화할 수 있다.

Fig. 1의 DCT RRU-Net에서 입력은 RGB 스트림과 DCT 스트림 두 가지로 구성된다. RGB 스트림은 RRU-Net의 Down Sampling 경로를 통해 시각 콘텐츠 및 이미지 수집 아티팩트를 탐색한다. DCT 스트림은 JALM의 입력이다. 학습 모듈의 각 Down Sampling 블록 출력은 Up Sampling 되어 Up Sampling 블록의 입력에 추가된다. RGB 스트림의 Down Sampling 블록을 통과한 산출물은 대응하는 Up Sampling 블록의 입력에 Skip Connection으로 Identity Mapping이 이뤄진다. Out Conv는 전처리를 거친 이미지와 같은 해상도로 1개 채널을 출력하며 픽셀 단위로 위조 영역의 확률을 구한다.

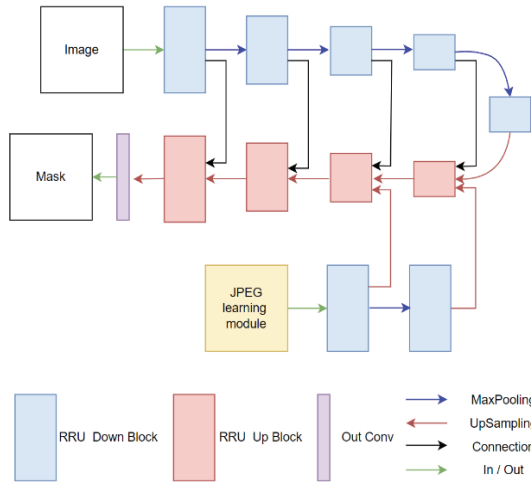


Fig. 1. DCT RRU-Net architecture.

DCT RRU-Net의 Down Sampling 블록과 Up Sampling 블록은 RRU-Net과 동일하다. JALM의 출력이 첫 번째와 두 번째 Up Sampling 블록에 연결될 때, 기존 입력과 결합된다. 이 과정에서 마지막 Down Sampling 출력의 Up Sampling과 같은 방식으로 해상도를 맞춰 입력에 더해진다. 하지만 이 방식은 시각 콘텐츠 및 이미지 수집 아티팩트를 탐색하기 위해 연구된 방법으로 DCT 도메인에서도 적절한지

검증이 필요하다. 본 연구에서 JALM을 Ringed Residual 블록으로 안정적인 학습 방법을 도출하기 위해 활성화 함수를 설계하였다.

$$y_f = F(x, W_i) + W_s \times x \quad (1)$$

$$y_b = (s(G(y_f) + 1) \times x) \quad (2)$$

수식 2에서 함수 s 는 Sigmoid, 수식 1의 $F = W_2\sigma(W_1 * x)$ 이며, σ 는 Relu 이다. Ringed Residual 구조에서 Sigmoid는 이미지 본질 속성의 차이를 증폭하기 위해 사용되었지만, DCT 도메인에서도 적절한 활성화 함수로 작용 가능한지 검증이 필요하다. 따라서 Leaky Relu를 사용하여 비교하였다. Leaky Relu는 Relu에 비해 연산속도가 느리지만, 음수 출력을 반영하며 DCT 스트림에 내재된 표현을 학습하기 적합하다고 판단했다. 음수 출력 결과를 유지하기 위해 JALM과 연결된 Down Sampling 블록의 Ringed Residual 구조의 함수 s , σ 를 Leaky Relu를 의미하는 l 로 재정의한다. 수정된 수식은 다음과 같다.

$$l = \begin{cases} x, & \text{if } x \geq 0 \\ x \times 0.01, & \text{otherwise} \end{cases} \quad (3)$$

$$F = W_2 l(W_1 \times x) \quad (4)$$

$$y_f = F(x, W_i) + W_s \times x \quad (5)$$

$$y_b = (l(G(y_f) + 1) \times x) \quad (6)$$

위 수식으로 구현된 DCT RRU-Net의 JALM Down Sampling 블록들은 RRU-Net과 같은 순서로 Ringed Residual 구조를 이룬다.

2.2 RRU-Net 과 DCT RRU-Net 의 비교

RRU-Net은 시각적 아티팩트 탐색으로 스플라이싱 위조에 대한 법의학적 단서를 발견한다. 또한, DRRU-Net은 압축 아티팩트를 탐색하고 압축과정에서 발생하는 위조에 대한 법의학적 단서를 추가로 발견할 수 있다. 단, DRRU-Net은 DCT 계수를 구하기 위해 입력 해상도는 가로 길이와 세로의 길이는 8의 배수로 고정되어야 한다. CNN의 특성상 각기 다른 해상도의 이미지를 학습에 사용하더라도 네트워크의 입력에는 고정된 해상도의 이미지가 요구된다. RRU-Net은 CASIA[18]와 COLUMB[20] 데이터셋을 훈련 및 평가에 사용했다. 해당 데이터셋에 대한 학습률을 높이기 위해 RRU-Net 입력의 해상도는 384x256으로 고정했다. 이는 CAISA의 평균적인 이미지 해상도이며 해상도가 큰 다른 데이터셋을 평가하기에 적절치 않다. 향후, 더 다양한 데이터셋의 평가를 위해 DCT RRU-Net의

입력 해상도는 512x512를 기준으로 한다. DCT RRU-Net의 고정된 입력을 위해 그리드 정렬 자르기 방식[14]을 사용한다. RRU-Net은 입력을 위한 이미지 전처리 작업에 리사이즈(Resize)와 자르기(Crop)를 사용했지만, 리사이즈는 이미지의 해상도를 줄이거나 늘리는 과정에서 이미지 본질 속성에 손상을 가할 수 있다. 따라서 본 논문의 실험에서는 이미지의 본질 속성을 유지하여 학습하기 위해 RRU-Net의 RGB 스트림 입력도 그리드 정렬 자르기 방식을 사용했다.

3. 실험 및 결과

3.1 실험 조건

DCT RRU-Net의 성능 평가를 위한 실험에서 DCT RRU-Net과 RRU-Net의 RGB 스트림을 학습하기 위한 Down Sampling 구조는 같다. 따라서 DCT RRU-Net은 RRU-Net에서 RGB 스트림을 학습하여 전이 학습이 가능하다. RRU-Net의 훈련과정에서 검증 점수가 증가 후 감소하기 이전 세대의 가중치를 저장한다. 이후, DCT RRU-Net은 해당 가중치의 Down Sampling 경로에 해당하는 블록들의 가중치들로 초기화하여 동결한다. 이후 JALM과 연결된 Down Sampling 경로와 Up Sampling 경로의 블록들만 훈련하여 RRU-Net 그리고 전이 학습을 하지 않은 DCT RRU-Net과 시멘틱 세그멘테이션 성능을 비교한다. 결과의 비교를 통해 RGB 도메인에서 사전 훈련에 대한 효과를 설명한다.

실험에 사용된 GPU는 NVIDIA GTX 1660이며 DRRU-Net은 Pytorch로 구현하였다. 성능 비교에 사용된 모든 모델은 DEFACTO의 Splicing 세트 이미지 1만 장을 학습하고, 평가를 위한 테스트 데이터세트는 훈련 및 검증 세트로 사용하지 않은 DEFACTO의 Splicing 세트 이미지 1만 장과 DEFACTO의 Copy-Move 세트 1만 장, CASIAv2 데이터세트의 위조된 이미지 세트 3000장을 사용한다.

DEFACTO의 데이터세트는 TIFF 포맷의 이미지만 포함한다. CASIAv2는 JPEG 파일이 아닌 TIFF, PNG 포맷의 이미지도 섞여서 저장되어있다. 딥러닝 네트워크상에서 압축 아티팩트를 수집하기 위해 훈련 데이터세트와 테스트 데이터세트의 JPEG 포맷이 아닌 이미지들은 Quality를 100으로 JPEG 압축이 수행된다.

네트워크를 통과하기 위한 RGB 스트림 입력을 위해 이미지의 정규화가 필요하다. [13]에서 이미지 X에 대하여 사용하는 정규화 방식은 $X / 255$ 이고, [14]에서 $(X - 127.5) / 127.5$ 로 이미지 픽셀값을 각 0~1, -1~1로 정규화한다.

DCT RRU-Net의 손실 함수(Loss Function)는 [13]의 실험과 마찬가지로 이진 교차 엔트로피(Binary Cross Entropy)를 사용하여 이미지의 픽셀 단위로 손실을 구한다. 옵티마이저

(Optimizer)는 Adam[18]을 학습률(Learning Rate) 0.001로 사용한다. 배치 사이즈(Batch Size)는 1이고, 훈련에 참여하지 않은 이미지 764장을 검증 데이터로 사용하여 훈련 루프 중에 검증 점수가 더 이상 높아지지 않고 6 에포크가 지나면 훈련을 멈춘 뒤, 검증 점수가 가장 높은 에포크의 모델을 저장하고 평가에 사용한다. 검증 점수는 Dice Coefficient로 계산한다. Dice Coefficient는 정답과 예측 영역의 겹치는 정도를 구하는 집합 연산이다. 혼동 행렬로 수식을 계산할 수 있지만 Pytorch의 라이브러리를 이용하여 혼동 행렬 없이 구할 수 있으며 미분 가능하여 손실 함수로 사용할 수 있다.

Dice Coefficient와 F1-Score는 의미는 다르지만 같은 수식으로 구할 수 있다. 이미지 위조 영역을 탐지하는 작업은 픽셀별로 위조(P)와 정상(N)으로 이진 분류할 수 있으므로 픽셀 단위로 정확도(Accuracy)를 계산할 수 있다. 하지만 데이터세트의 각 이미지 내의 위조 영역 픽셀 수와 정상 영역의 픽셀 수의 비율은 고정되어 있지 않고 불균형하므로 정확도만으로 위조 픽셀 검출 성능에 대한 평가가 올바르게 이뤄지지 않을 수 있다. 예를 들어 100x100 해상도의 이미지에서 30개 픽셀만 위조되었다는 전제하에 훈련이 제대로 이뤄지지 않아서 모든 이미지의 픽셀을 모두 음성(정상)으로 예측하면 정확도는 0.997로 뛰어난 모델처럼 보인다. 따라서 양성 픽셀과 음성 픽셀의 불균형한 분포에서 신뢰할 수 있는 분류의 평가를 위해 혼동 행렬에서 구한 F1-Score는 Dice로 표기하여 사용했다.

$$DACC_p = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$PRE_p = \frac{TP}{TP+FP} \quad (8)$$

$$DREC_p = \frac{TP}{TP+FN} \quad (9)$$

$$F1_p = \frac{2(PRE_p \times DREC_p)}{PRE_p + DREC_p} = \frac{2TP}{2TP+FP+FN} = Dice \quad (10)$$

수식 7에서 첨자 p는 pixel을 기준으로 수식을 적용했음을 의미한다. TP(True Positive)는 위조된 픽셀의 옳은 탐지에 해당하고, TN(True Negative)은 정상 픽셀의 옳은 탐지, FP(False Positive)는 정상 픽셀을 위조로 오 탐지, FN(False Negative)은 위조된 픽셀의 미탐지에 해당한다. 각 이미지에서 구한 Dice Coefficient는 모두 더하여 평균값을 취한다. 모든 결과값은 넷째 자리에서 버림을 취하며 셋째 자리까지 사용한다.

3.2 실험 결과

[14]의 JALM 사전 훈련 방법과 다른 DCT RRU-Net의 사

전 훈련 방식을 설명하고 RRU-Net, 사전 훈련을 하지 않은 DCT RRU-Net과 성능 차이를 검증한다. [14]에서는 DCT 스트림을 이용하여 JALM을 사전 훈련한다. 이는 HRNet의 위조 도메인에서 RGB 스트림의 사전 훈련은 위조 영역 탐지에 유의미한 작업이 아니기 때문일 것이다.

HRNet과 마찬가지로 일반적인 CNN과 JALM의 결합은 DCT 도메인에 의존하여 위조를 탐지할 것이다. RGB 스트림을 이용하여 이미지 수집 아티팩트에서 위조 영역의 속성차이를 탐지할 수 있는 RRU-Net은 JALM과 결합했을 때, DCT 도메인에 의존하여 가중치가 편향되어 최적화될 가능성을 염려해야 한다. DCT RRU-Net의 가중치들을 한번에 학습시키면 위조 영역의 단서를 상대적으로 쉽게 찾을 수 있는 DCT 스트림에 의존하도록 학습이 이뤄질 것이다. 이를 방지하기 위해 RGB 스트림을 학습하는 RRU-Net을 훈련하고, 검증 점수가 더 이상 증가되지 않는 에포크에서 훈련을 중단한다. 이 RRU-Net의 Down Sampling 구간의 가중치를 DCT RRU-Net에 전이 학습한다. 해당 가중치는 훈련을 통해 가중치 갱신이 이뤄지지 않고 동결되며 DCT RRU-Net은 Up Sampling 구간과 JALM과 연결된 Down Sampling 경로의 레이어만 학습이 이루어진다. 이 작업으로 RGB 스트림에서도 의미 있는 특징을 추출할 수 있으며 JALM은 포함된 레이어의 가중치를 동결하지 않으므로 가중치를 동결한 상태보다 더 높은 특징 추출 성능을 보일 수 있다.

본 실험에서 RRU-Net과 DCT RRU-Net, DCT RRU-Net TL을 비교한다. RRU-Net의 1 에포크에 소요한 시간은 훈련 및 검증 작업을 포함하여 평균 7,335초이다. 19 에포크까지 학습이 진행된 후 6 에포크 동안 검증 점수가 오르지 않았다. 최종적으로 검증 점수는 0.681이고, 마지막 훈련 손실(Train Loss)은 0.308이다. DCT RRU-Net의 1 에포크에 소요한 시간은 9,780초이며, 6 에포크 이후로 6 에포크 동안 검증 점수가 오르지 않았다. 최종적으로 검증 점수는 0.903, 마지막 훈련 손실은 0.090이다. RRU-Net과 비교하여 1 에포크 당 소요 시간은 더 길지만 더 적은 에포크에서 훈련이 종료됐으며, 낮은 훈련 손실을 기록했고 검증 점수 또한 더 높게 나타났다. DCT RRU-Net TL은 위의 훈련된 RRU-Net의 Down Sampling 블록들의 가중치를 전이 학습하여 동결한 모델이다. 1 에포크에 소요한 시간은 평균 8,997초다. 검증 점수는 6 에포크 이후로 6 에포크 동안 0.901에서 더 오르지 않았고, 마지막 훈련 손실은 0.090이다. 처음부터 훈련된 DCT RRU-Net과 TL 모델의 검증 점수와 손실이 유사하지만 테스트에서 DCT RRU-Net TL은 모든 실험을 통틀어 가장 높은 탐지율 및 다른 데이터셋에 대한 일반화 성능을 보였다.

Table 1. 픽셀 수준에서 Splicing 및 Copy-Move 위조 감지 성능 비교

Method		DEFACTO Splicing	DEFACTO Copy-Move	CASIAv2
RRU-Net	Dice	0.306	0.062	0.081
	Acc	0.988	0.963	0.914
DCT RRU-Net	Dice	0.830	0.652	0.547
	Acc	0.997	0.984	0.864
DCT RRU-Net TL	Dice	0.830	0.716	0.830
	Acc	0.997	0.992	0.967





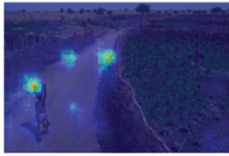
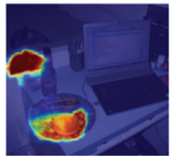
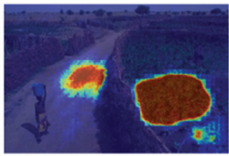

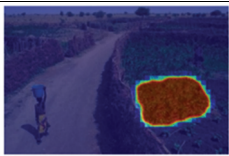
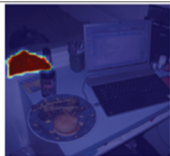
RRU-Net의 Dice Coefficient는 DEFACTO Splicing 세트에서 0.306, DEFACTO Copy-Move에서 0.062 CASIAv2에서 0.081이다. 픽셀 단위 정확도는 각 데이터셋에서 0.988, 0.963, 0.914이다. 예상보다 점수가 낮게 나온 이유는 [13]의 훈련 및 평가 방식의 차이이다. CASIA[17]를 훈련 및 평가에 사용한 [13]의 실험에서 RRU-Net을 더 잘 학습시키기 위해 훈련 데이터셋을 압축 및 무작위 뒤집기 기법으로 증대시켰다. CASIA와 COLUMB는 위조 이미지의 원본을 포함하고 있다. 훈련을 위한 이미지는 무작위로 선별되었으므로 훈련 또는 평가에 위조된 이미지에 대한 원본 이미지가 사용되었을 가능성이 크다. 또한 테스트 데이터셋의 일반 스플라이싱 위조 이미지보다 해당 데이터셋을 이용하여 임의로 생성한 JPEG 압축과 노이즈를 가한 이미지를 더 많이 사용하여 스플라이싱 위조 탐지에 대한 명확한 성능지표로 보기 어렵다. RRU-Net의 픽셀 단위 정확도가 높은 이유는 이미지상에 위조된 영역보다 그렇지 않은 영역이 대부분인 상황에서 RRU-Net이 위조 영역을 찾지 못하여 이미지의 대부분 픽셀에 대해 정상 영역으로 분류했기 때문이며, 이는 Dice 값과 Fig. 2를 통해 확인할 수 있다.

제안하는 DCT RRU-Net은 랜덤으로 가중치를 초기화하고 훈련했을 때, Dice Coefficient는 DEFACTO Splicing 세트에서 0.830 DEFACTO Copy-Move 세트에서 0.652, CASIAv2에서 0.547이다. 정확도는 각 데이터셋에서 0.997, 0.984, 0.864이며, 이는 RRU-Net의 결과보다는 우수하지만 만족할만한 결과는 아니다.

Table 2는 CASIAv2, DEFACTO Copy-Move 각 세트에서 Copy-Move 위조 이미지를 한 장씩 뽑아 RRU-Net, DCT RRU-Net, DCT RRU-Net TL 모델로 예측한 결과이다. RRU-Net은 Table 2에서 보이는 것과 같이, 테스트 데이터셋의 대부분 이미지에서 위조 영역을 정확하게 식별하지 못했

다. DRRU-Net은 훈련 데이터셋과 다른 처리 과정을 거친 CASIAv2에서 불안정한 예측을 보인다. 위조 영역에 대한 예측을 포함하지만 정상 영역까지 포함하는 이미지의 빈도가 높았다. DCT RRU-Net TL은 다른 네트워크와 비교하여 상대적으로 정확하게 위조 영역을 탐지했다.

Table 2. RRU-Net, DRRU-Net, TL 모델 세그멘테이션 성능 비교

	Sample 1	Sample 2
Input		
Ground Truth		
RRU-Net		
DCT RRU-Net		
DCT RRU-Net TL		

본 실험에서 RRU-Net은 저조한 성능을 보였지만 저조한 성능의 RRU-Net Down Sampling 구간의 가중치를 전이 학습한 DCT RRU-Net TL은 비교한 네트워크 중 가장 높은 성능을 나타냈다. DCT RRU-Net TL과 비교해서 DCT RRU-Net은 CAT-Net과 마찬가지로 RGB 스트림보다 DCT 스트림을 위주로 탐색하도록 가중치가 최적화된 것으로 볼 수 있다. 따라서 RGB 스트림에서만 발견할 수 있는 단서를 찾는 RRU-Net을 전이 학습한 DCT RRU-Net TL은 DCT RRU-Net처럼 DCT 스트림에 의존하지 않고 상대적으로 RGB 스트림과 DCT 스트림에서 균형 있는 탐색이 이루어졌음을 알 수 있다.

4. 결론

CAT-Net의 JALM을 RRU-Net의 Ringed Residual 구조에 효과적으로 적용하는 DCT RRU-Net과 쉽고 간단한 사전 훈련 방법을 제시했다. 실험을 통해 [5]에서 제시한 다른 데이터셋을 이용해야 하는 JALM의 사전 훈련 방법을 사용하지 않고, 간단하게 RGB 스트림을 사전 훈련하면 같은 데이터셋 이미지를 사용하면서도 유효하게 성능을 올릴 수 있음을 보였다. 향후 연구에서 RRU-Net은 RGB 스트림으로 학습하여 다른 버전의 확장된 DCT RRU-Net 구조에 성능 향상을 위한 모듈로 사용할 수 있다. 위조에 대한 탐지 작업은 DCT RRU-Net은 CAT-Net, RRU-Net과 비교하여 사전 작업 없이 동등한 조건에서 학습했을 때, 다른 위조 도메인(Copy-Move)의 데이터셋에서도 가장 높은 일반화 성능을 보였고 특히, 훈련 데이터셋과 다른 위조 도메인을 포함하고, 위조 처리 과정도 다른 CAISAv2에서도 가장 높은 성능을 보였다. 본 연구 여건으로 [5]의 실험처럼 많은 데이터셋을 훈련하지 못했지만 [5]와 같이 충분한 데이터셋을 훈련한 DRRU-Net은 이미지 Splicing 위조 영역 세그멘테이션 모델 중, 가장 우수한 성능을 보일 것으로 기대된다.

참고문헌

1. D. L. M. Sacchi, F. Agnoli, and E. F. Loftus, "Changing history: doctored photographs affect memory for past public events," vol. 21, no. 8, pp. 1005–1022, Dec. 2007, doi: 10.1002/acp.1394.
2. M. Mishra and F. L. D. M. C. Adhikary, "Digital Image Tamper Detection Techniques - A Comprehensive Study," Jun. 2013, doi: 10.48550/arxiv.1306.6737.
3. C. N. Bharti and P. Tandel, "A survey of image forgery detection techniques," 2016, pp. 877–881, doi: 10.1109/WiSPNET.2016.7566257
4. W. N. Nathalie Diane, S. Xingming, and F. K. Moise, "A Survey of Partition-Based Techniques for Copy-Move Forgery Detection," vol. 2014, pp. 975456–13, Jul. 2014, doi: 10.1155/2014/975456.
5. M. D. Ansari, S. P. Ghrera, and V. Tyagi, "Pixel-Based Image Forgery Detection: A Review," vol. 55, no. 1, pp. 40–46, Jan. 2014, doi: 10.1080/09747338.2014.921415.
6. G. K. Birajdar and V. H. Mankar, "Digital image forgery detection using passive techniques: A survey," vol. 10, no. 3, pp. 226–245, Oct. 2013, doi: 10.1016/j.diin.2013.04.007.
7. T. Qazi et al., "Survey on blind image forgery detection," vol. 7, no. 7, pp. 660–670, Oct. 2013, doi: 10.1049/iet-ipt.2012.0388.

8. Y. Shi, C. Chen, and W. Chen, "A natural image model approach to splicing detection," 2007, pp. 51–62, doi: 10.1145/1288869.1288878
9. Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on Markov features in DCT and DWT domain," vol. 45, no. 12, pp. 4292–4299, Dec. 2012, doi: 10.1016/j.patcog.2012.05.014.
10. S. Velliangiri and J. Premalatha, "A Novel Forgery Detection in Image Frames of the Videos Using Enhanced Convolutional Neural Network in Face Images," vol. 125, no. 2, pp. 625–645, Nov. 2020, doi: 10.32604/cmcs.2020.010869.
11. H. Mo, B. Chen, and W. Luo, "Fake Faces Identification via Convolutional Neural Network," 2018, pp. 43–47, doi: 10.1145/3206004.3206009
12. Y. Wu, W. AbdAlmageed, and P. Natarajan, "ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries With Anomalous Features," 2019, pp. 9535–9544, doi: 10.1109/CVPR.2019.00977
13. X. Bi, Y. Wei, B. Xiao, and W. Li, "RRU-Net: The Ringed Residual U-Net for Image Splicing Forgery Detection," 2019, pp. 30–39, doi: 10.1109/CVPRW.2019.00010
14. M.-J. Kwon, S.-H. Nam, I.-J. Yu, H.-K. Lee, and C. Kim, "Learning JPEG Compression Artifacts for Image Manipulation Detection and Localization," vol. 130, no. 8, pp. 1875–1895, 2022, doi: 10.1007/s11263-022-01617-5.
15. J. Wang et al., "Deep High-Resolution Representation Learning for Visual Recognition," vol. 43, no. 10, pp. 3349–3364, Oct. 2021, doi: 10.1109/TPAMI.2020.2983686. org.libproxy.smu.ac.kr/document/9052469
16. Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848
17. Jing Dong, Wei Wang, and Tieniu Tan, "CASIA Image Tampering Detection Evaluation Database," 2013, pp. 422–426, doi: 10.1109/ChinaSIP.2013.6625374
18. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, doi: 10.48550/ARXIV.1412.6980.

접수일: 2023년 1월 26일, 심사일: 2023년 3월 7일,
게재확정일: 2023년 3월 20일