캡스톤 디자인
'딥페이크 탐지'

#2. 특허, 개발물, 논문 정보 조사

김지수, 김민지, 민지민

특허

‘Method And Systems For Detecting Deepfakes’

- 등록일: 2021. 01. 22
- 출원인: ZeroFOX, Inc(미국)

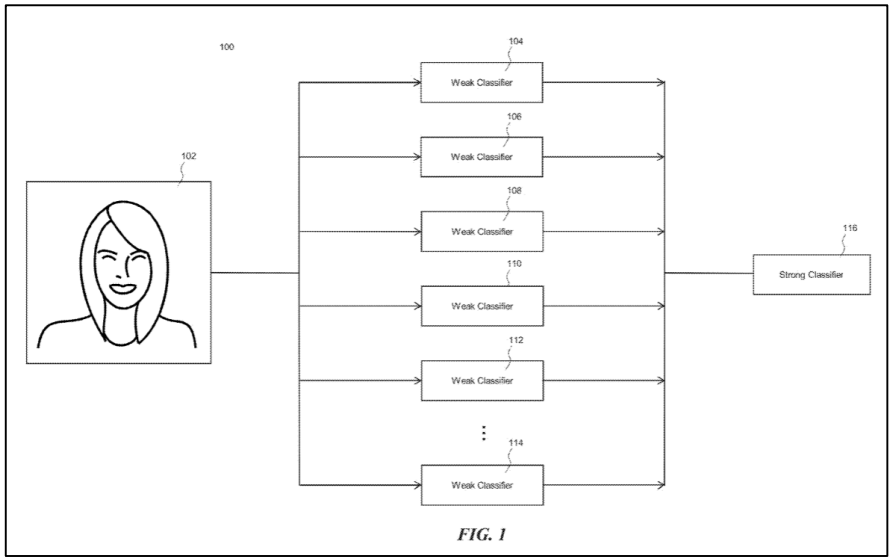


FIG. 1

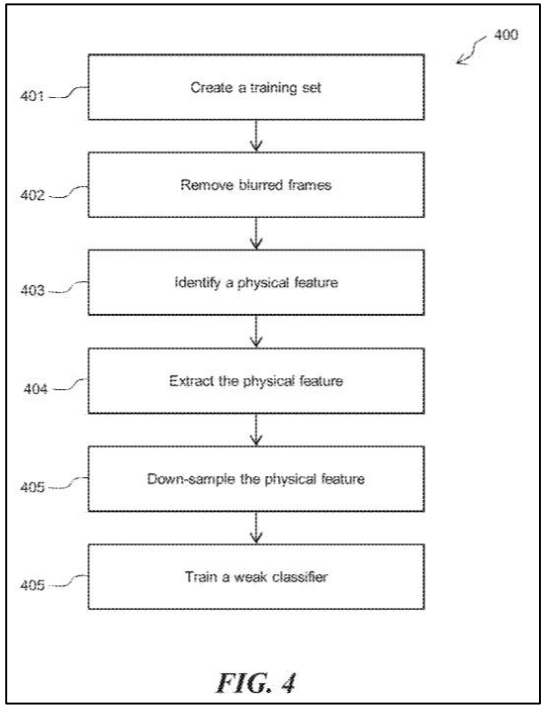


FIG. 4

- **요약**

다수의 약한 분류기로부터 **예측 결과를 서버가 수신**하고, 그 결과를 강한 분류기로 보내고, 강한 분류기는 **예측 결과를 분석하고 합성영상 여부를 판단**한다.

특허

‘Method And Systems For Detecting Deepfakes’

- **배경**

  - 딥페이크 탐지를 위해 시도된 거의 대부분의 기술은 네트워크가 직접 분류하는 방법인 강한 분류기를 포함한다. 그와 대조되는 분류기로 특징을 분석하고 탐지하는 신경망인 약한 분류기가 있다. 그러나 그들은 그 자체로는 분류를 느슨하게 예측.

- **해결 수단**

  1) 합성 영상물 중 흐릿한 프레임 제거

  2) 첫번째 약한 분류자는 입을 탐지하도록 훈련되고, 치아나 얼굴의 털과 관련된 입의 불규칙성을 감지해서 그를 기반으로 첫번째 예측 결과 생성

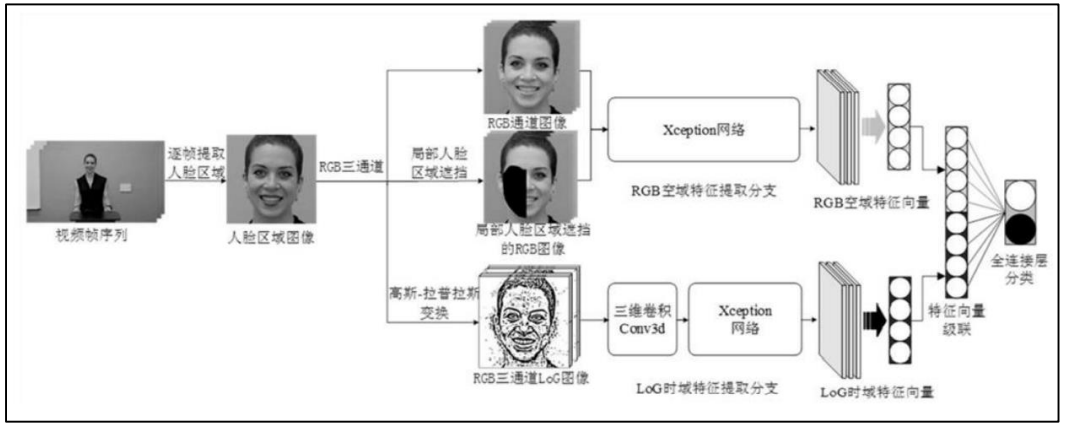
  3) 두번째 약한 분류자는 머리의 움직임을 감지하고, 그를 기반으로 펄스를 계산하며 펄스의 불규칙성을 기반으로 두번째 예측 결과 생성

  4) 위와 같이 다양한 측면으로의 예측 결과를 생성하여 서버를 통해 강한 분류기로 예측 결과를 보내고, 합성 여부를 결정

특허

‘RGB 공간 영역 특징과 LoG 시간 도메인 특징을 결합시킨 Deepfake 영상 검파 방법과 시스템’

- 등록일: 2021. 03. 01
- 출원인: South China University of Technology(중국)



- **요약**

얼굴 합성 영상에서 **RGB공역 특징과 LoG시간 영역 특징을 추출**하여 **국부적인 얼굴 영역 차단 처리를 결합**함으로써 모델의 합성영상 검출 능력을 향상시킨다.

특허

‘RGB 공간 영역 특징과 LoG 시간 도메인 특징을 결합시킨 Deepfake 영상 검파 방법과 시스템’

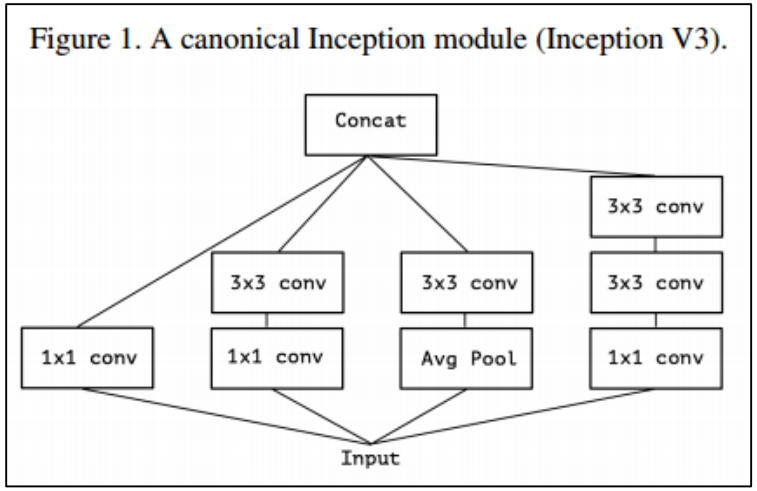
- **해결 수단**

  1) 영상 프레임에서 얼굴 영역 RGB 3채널 이미지 추출

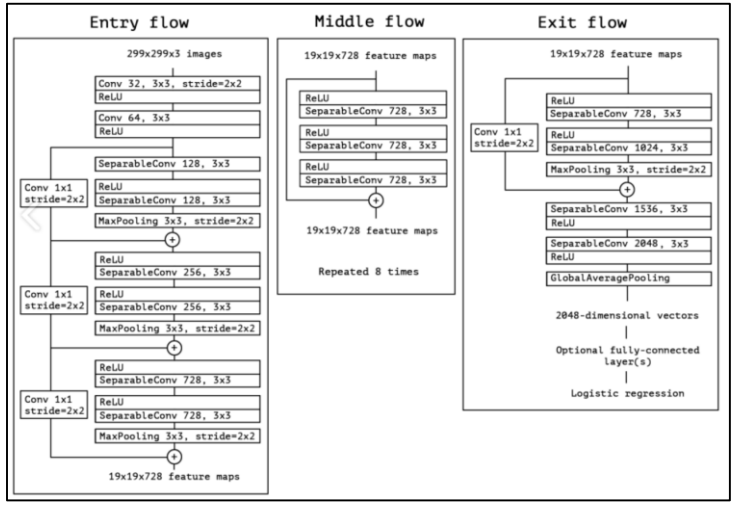
  2) 1)의 이미지와 얼굴의 부분 영역을 가린 이미지를 Xception에 입력하여 RGB 특징 추출

  3) 연속 프레임의 RGB 이미지를 LoG 변환해 3차원 롤링과 Xception을 거쳐 합성 영상 검출 능력 향상

<Inception>



<Xception>



Xception이란?

- **Inception 모듈**에 대한 고찰로 탄생한 모델

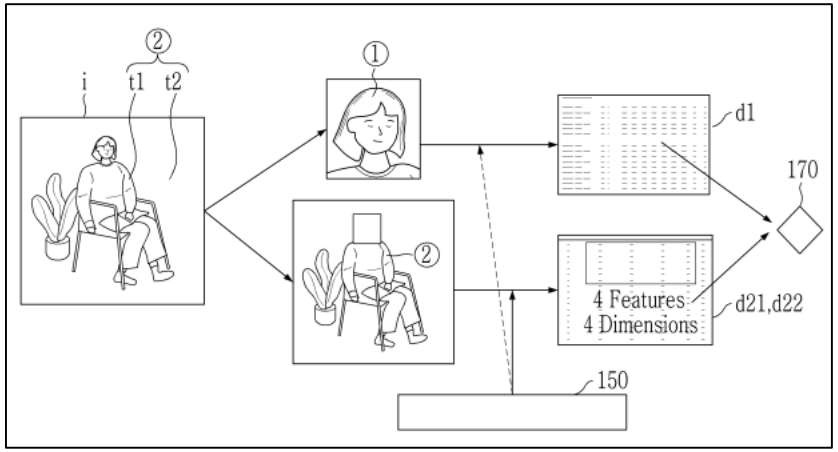
- Inception 모듈의 1x1 conv는 **cross-channel correlation**을 계산하고, 3x3 conv은 **spatial correlation**을 수행하는데, 이를 잘 **분해해서 계산**했기 때문에 Inception 모듈의 성능이 좋음

- **Xception**: Inception 모듈의 cross-channel correlation과 spatial correlation을 **완벽히 독립적으로 계산**하기 위한 고안된 모델

특허

‘페이크 얼굴 검출 장치 및 검출 방법’

- 등록일: 2020년 5월 21일
- 출원인: 경일대학교 산합협력단



- **요약**

- **얼굴 영역과, 얼굴이 아닌 영역을 분리**한 후 각 영역의 **특징점을 추출**하고 그 특징점들의 비교를 통해 **위조 여부를 판별**

특허

‘페이크 얼굴 검출 장치 및 검출 방법’

- **해결 수단**

  1) 입력된 얼굴 영상의 각 픽셀을 **밝기 값에 대해 푸리에 변환**(주파수 도메인 변환)

  2) 변환된 값의 절대값을 **로그 스케일**로 나타낸 후 **저주파 성분이 중심**에 오도록 위치 변환

  3) 변환된 값을 다수의 **동심원으로 구획**하고, 각 **동심원의 평균 에너지** 값을 산출해 **제1 1차원 특징 벡터**를 생성

  4) 입력된 얼굴 영상으로부터 **로컬바이너리패턴 코드를 계산**한 코드를 **히스토그램화**

  5) 히스토그램의 합이 1이 되도록 **정규화**하여 **제2 1차원 특징 벡터**를 생성

  6) **서포트벡터머신**을 이용해 산출된 두 특징 벡터를 **각 기준 얼굴 특징 벡터와 비교**

  7) 두 판정 결과를 조합해 **위조 여부를 최종으로 판정**

푸리에 변환(Fourier Transform)이란?

$$F(u)=\int_{-\infty}^{\infty} f(x)e^{-j2\pi ux}dx$$

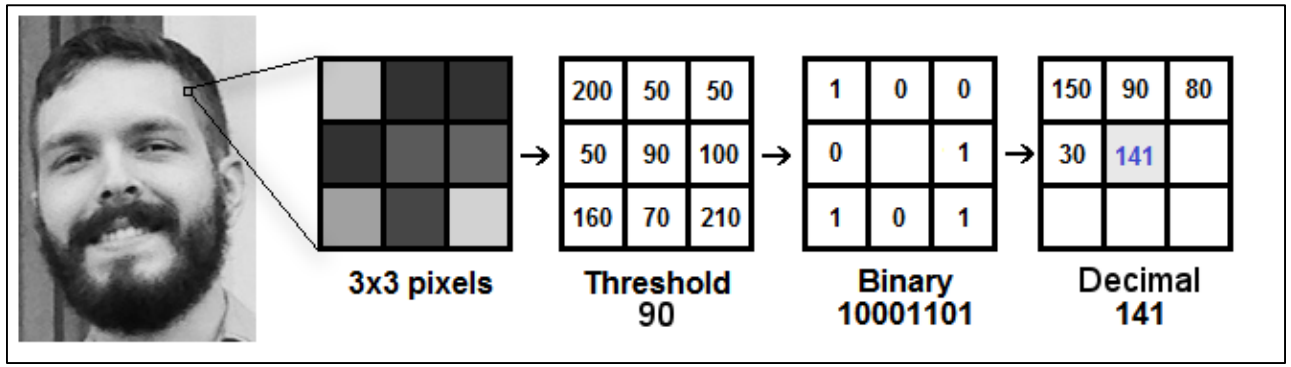
- 임의의 **입력 신호**를 다양한 **주파수를 갖는 주기함수들의 합**으로 분해하여 표현하는 것

- 컴퓨터 비전에는 **‘spatial domain’**에서 **‘frequency domain’**으로의 변환이라고 표현

- 왜 주파수 도메인으로 변환하는지?

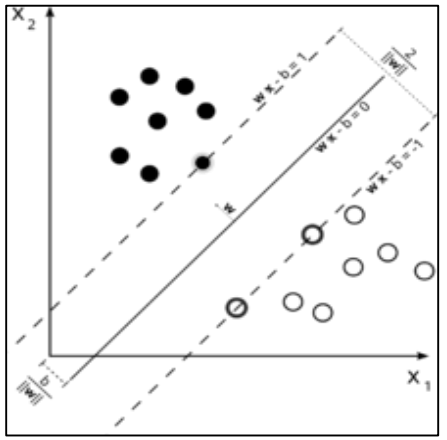
로컬 바이너리 패턴(LBP)이란?

- Local Binary Pattern(LBP) 알고리즘은 말그대로 **지역적인 이진 패턴을 계산**



- 어떤 **주변의 값을 2진수로 표현**한 뒤, 계산
  - 셀 중심 픽셀과 이웃하는 픽셀의 크기를 비교한 (크면 1 작으면 0) 값을 나열해 이진값을 얻음
  - 모든 픽셀에 대해 이진값을 계산한 후 히스토그램화 한다
  - **영상의 texture를 숫자로 표현하는 기능**

서포트 벡터 머신(SVM)이란?

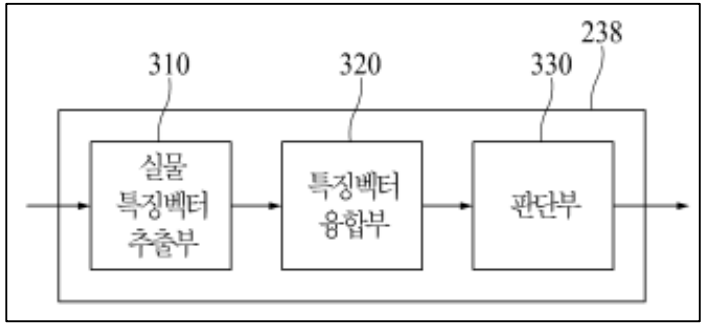


- 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 카테고리에 속할지 판단하는 **이진 선형 분류 모델을 만듦**
- 만들어진 분류 모델은 **데이터가 사상된 공간에서 경계로 표현**
- SVM알고리즘은 여러 모델들을 기반으로 **가장 큰 폭을 가진 경계를 찾는 알고리즘**

특허

‘딥러닝 기반의 얼굴 인식 모델을 이용하여 실물 이미지를 판단할 수 있는 안면인식시스템’

- 등록일: 2019년 12월 3일
- 출원인: 주식회사 포스코아이씨티



- **배경**

- 일반적인 안면인식시스템은 **등록된 사용자의 얼굴이 포함된 사진으로 인증**을 수행하는 경우를 **식별할 수 없는 문제점**이 있다. 또한 이러한 문제를 해결하기 위해 **별도의 특수카메라**를 이용하여 이를 식별 가능하지만, **비용이 많이 소모**된다는 문제가 있다.

- **요약**

- 별도의 특수 카메라 없이 얼굴 이미지의 **깊이, 빛 반사, RGB**를 표현하는 벡터들을 **융합한 융합벡터를 사용**해 입력된 이미지가 실물이미지인지 여부를 판단한다.

특허

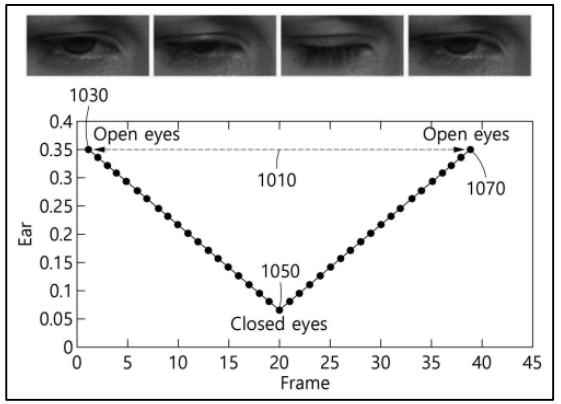
‘딥러닝 기반의 얼굴 인식 모델을 이용하여 실물 이미지를 판단할 수 있는 안면인식시스템’

- **해결 수단**
  - 입력된 얼굴 이미지로부터 **깊이 특징벡터** 및 이미지의 빛 반사를 표현하는 **반사 특징벡터**와 이미지의 **RGB 특징벡터**를 추출하는 **실물 특징벡터 추출부**
  - 깊이 특징벡터와 반사 특징벡터 중 **적어도 하나를 RGB 특징벡터와 융합**한 융합 특징벡터를 생성하는 **특징벡터 융합부**
  - 융합 특징벡터를 이용해 얼굴 이미지가 사람을 촬영한 **실물 이미지인지 여부를 판단**하는 판단부를 통해 최종 판단

특허

‘딥러닝을 통해 생성된 가짜 동영상의 무결성 검증 방법 및 시스템’

- 등록일: 2019년 12월 18일
- 출원인: 건국대학교 산학협력단



- **배경**

- 최근 GAN 모델 기반의 **generator 기술이 고도로 발전**함에 따라 단순히 픽셀을 기반으로 무결성을 검증하는 방법의 활용성이 떨어졌다.

- **요약**

- 동영상에 포함된 인물의 **눈 깜빡임에 기초**하여 가짜 동영상 여부를 결정

특허

‘딥러닝을 통해 생성된 가짜 동영상의 무결성 검증 방법 및 시스템’

- **해결 수단**
  - 입력된 동영상으로부터 인물의 **눈 깜빡임 횟수**, 인물의 **성별/나이에 대한 예측** 정보, **행동에 대한 예측** 정보 생성

  - 예측된 **성별 및 나이**에 따라 정의된 **기준 인물의 눈 깜빡임의 평균 횟수를 선택**하고, 예측된 **행동**에 따라 정의된 **기준 인물의 눈 깜빡임 횟수를 연산 가중치**로 두어 기준 횟수를 계산

  - 입력된 **영상의 인물의 눈 깜빡임 횟수와 기준 횟수의 차이**를 기준으로 **가짜 동영상인지 여부를 결정**

Mobilenet-face-extractor

- MobileNet은 Depthwise separable convolution을 활용해서 모델을 경량화했다.
- Xception은 Depthwise separable convolution을 활용해서 감소한 파라미터 수 만큼 층을 쌓아 성능을 높이는데 집중한 것에 반면 경량화에 집중한다.
- Depthwise Separable Convolution은 각 입력 채널에 대해 3*3 conv 하나의 필터가 연산을 수행하여 하나의 feature 맵을 생성
- Pointwise convolution은 Depthwise convolution이 생성한 피쳐맵들을 1*1 conv 채널 수를 조절

Mobilenet-face-extractor

- 메모리가 제한된 환경에서 MobileNet을 최적으로 맞추기 위해 두 개의 파라미터를 latency와 accuracy를 조절하는 파리미터 중 하나는 두께를 결정
- 여기서 두께는 필터 수를 의미합니다.
- 다른 하나는 모델의 연산량을 감소시키기 위해 사용한다. 입력 이미지에 적용하여 해상도를 낮춘다.

- https://www.kaggle.com/unkownhihi/mobilenet-face-extractor-helper-code

3D CNN

- 비디오 프레임 간에 가능한 불일치 큐를 활용하는 것을 목표로 하고 딥페이크 비디오를 감지하기 위한 시간적 dropout 3차원 convolution신경망(TD-3DCNN)을 제안

- 이 접근 방식에서 비디오에서 샘플링 된 고정 길이 프레임 볼륨은 3차원 convolution 신경망(3DCNN)에 공급되어 다양한 규모의 특징을 추출하고 실제인지 가짜인지 식별

- 각 배치에서 프레임을 무작위로 샘플링하기 위해 시간적 dropout 작업이 도입 단순하지만 효과적인 데이터 증강의 역할을 하며 표현 및 일반화 능력을 향상시켜 모델 과적합을 피하고 탐지 정확도를 향상

https://www.kaggle.com/keremt/dfdc-3d-2d-inc-cutmix-with-3d-model-fix

Detecting Deepfakes with Metric Learning

- 게재된 저널 : **IEEE2020 8th International Workshop on Biometrics and Forensics (IWBF)**
- **발행일 : 2020년 4월 29일**
- **해결수단**

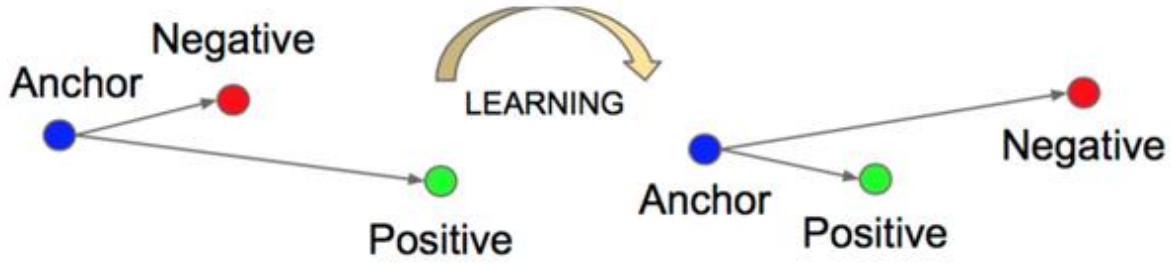
가짜 비디오와 실제 비디오 임베딩 벡터를 구분하기 위해 triplet network 를 사용한다.

1. MTCNN을 사용해서 프레임에서 얼굴을 추출한다.
2. facenet은 피처 공간의 각 면에 대해 512차원 임베딩을 생성한다
3. 각각의 얼굴이 피처 공간에서 작은 클러스터 차지
4. 온라인 틀리플렛 마이닝으로 semi hard triple을 생성한다
5. 세 triple을 이용하여 가짜 프레임과 진짜 프레임의 임베딩은 triplet loss를 통해 뚜렷하게 분리된다.

Triplet 네트워크는 유사한 기능이 함께 그룹화되고 다른 기능이 기능 공간에서 크게 떨어져 배치되는 학습 유형이다.

Triplet Loss

- Triplet loss는 무작위 데이터셋을 뽑은 후 positive pair와 negative를 샘플한 후, positive는 가까이, negative는 멀리 배치하는 방법
- Triplet mining은 triplet loss에 따라 dataset의 카테고리를 나눌 수 있다. 종류는 Easy, Hard, Semi-hard
- Easy triplet은 분리가 잘되어 있는 경우
- Hard triplet은 negative가 positive보다 anchor에 가까운 경우
- Semi-hard triple은 negative가 positive 보다 멀리 떨어져 있지만 충분한 margin을 넘지 못해 loss가 양수인 경우를 말한다.



MTD-Net: Learning to Detect Deepfakes Images by Multi-Scale Texture Difference

- **게재된 저널 : IEEE Transactions on Information Forensics and Security (Volume: 16)**
- **발행일 : 2021년 8월 3일**
- **<u>해결수단</u>**

    1. 텍스처 차별적 정보활용

    픽셀 강도 정보와 픽셀 기울기 정보를 사용하여 텍스처 차이 정보에 대한 고정된 설명을 제공

    2. 다단계 정보 추출

    확장 convolution을 사용하여 텍스처 차인 feature 모듈의 출력인 형상 맵에서 다른 스케일 형상 추출

MTD-Net: Learning to Detect Deepfakes Images by Multi-Scale Texture Difference

3. 얼굴 조작감지를 위한 네트워크

멀티 스케일 텍스처 차이 모델 MTD-Net 은 텍스처 차이 기능 모듈과 다중 스케일 정보 모듈로 구성

- 텍스처 차이기능 모듈은 바로가기 같은 하단 레이어 부근의 직관적인 특징과 상단 레이어 부근의 추상적인 특징을 결합할 수 있기때문에 ResNet-18로 구축됨 이 기능을 최대한 활용하기 위해 CDC로 대체된다
- ASPP 블록으로 구성된 다중 스케일 정보 모듈은 서로 다른 스케일 정보를 캡처하기 위해 입력 피쳐 맵에 서로 다른 확장 속도로 추출된 형상의 적응형 재보정을 학습하는데 1*1convolution 레이어가 사용됨
- 다중 스케일 정보 모듈 이후 글로벌 평균 풀링을 사용하여 공간 정보를 채널 통계로 압축하고 기능 정보를 완전히 연결된 계층으로 전송하여 최종 분류한다

MTD-Net은 실제 영역과 가짜 영역을 포함하는 더 큰 범위의 특징을 사용하여 얼굴 위조 탐지에 더 좋음

A Deepfakes detection technique based on two-stream network

- **게재된 저널 : Journal of Cyber Security (Volume 5, Issue 2, 2020)**
- **발행일 : 2020년 3월 10일**
- **<u>해결수단</u>**
  - 노이즈 특성을 이용하여 영상 압축 능력을 향상시킨다.
  - 노이즈 추출 영역은 딥러닝 방식으로 생성된 영역
  - 노이즈 스트림을 이용해 Deepfake 압축 특성 문제(서로 다른 압축률)를 해결하고, 서로 다른 복잡한 장면은 EfficientNet과 공동으로 향상 시킨다.
  - 영상을 프레임 시퀀스로 캡쳐한 후 얼굴 검출기를 이용하여 프레임 내 얼굴 정보를 추출

  - 모델 결정 계층에서 융합하여 두 스트림(상부, 하부)을 개별적으로 훈련한다
  - 하부는 노이즈 필터를 사용하여 얼굴의 소음 특성을 획득하여 EfficientNet으로 전송
  - 상부는 얼굴 정보 전체를 직접 훈련시켜서 진짜 얼굴 특징의 분포 차이를 학습
  - 두 결과를 융합하여 비디오의 모든 프레임을 예측하고 그 결과의 평균값으로 딥 페이크에 속하는지 확률로 예측

Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법

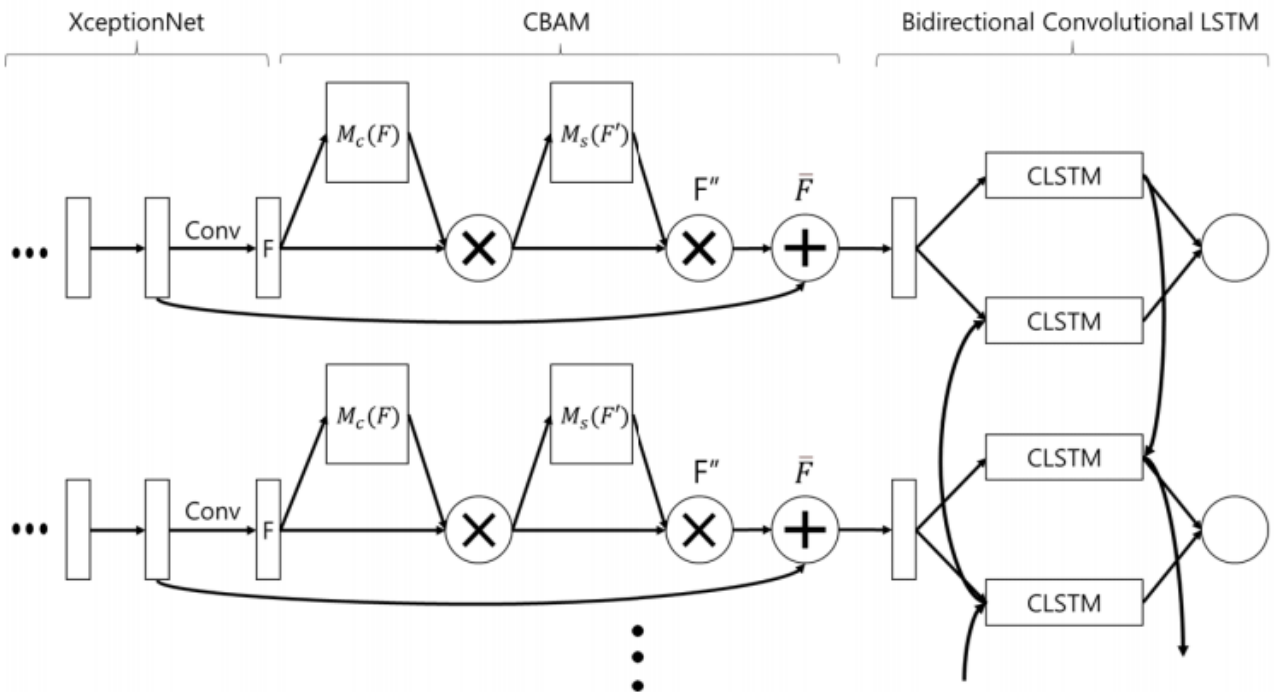
- 요즘 무분별한 가짜 동영상이 크게 증가하였으며 이는 개인 정보 침해, 가짜 뉴스, 사기 등에 문제로 이어질 수 있다. 따라서 사람의 눈으로도 진위를 가릴 수 없는 가짜 동영상을 탐지할 수 있는 방안이 필요하게 됨.
- 이에 본 논문에서는 **Bidirectional Convolutional LSTM과 어텐션 모듈(Attention module)을 적용한 딥페이크 탐지 모델을 제안** → 양방향 순환 신경망을 활용한 딥페이크 탐지 모델

Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법

- 본 논문에서 제안하는 모델은 **Convolution Block**, **Sequential layer**, **Fully-Connected layer**로 구성된다.
- Convolution Block에서는 프레임 단위로 분할된 동영상의 이미지 데이터로부터 **각 프레임의 특징을 추출**한다. Sequential layer는 **프레임 간 연결관계를 학습**한다. Fully-Connected layer에는 입력된 동영상이 딥페이크로 생성된 가짜 동영상인지 실제 동영상인지를 판별한다.

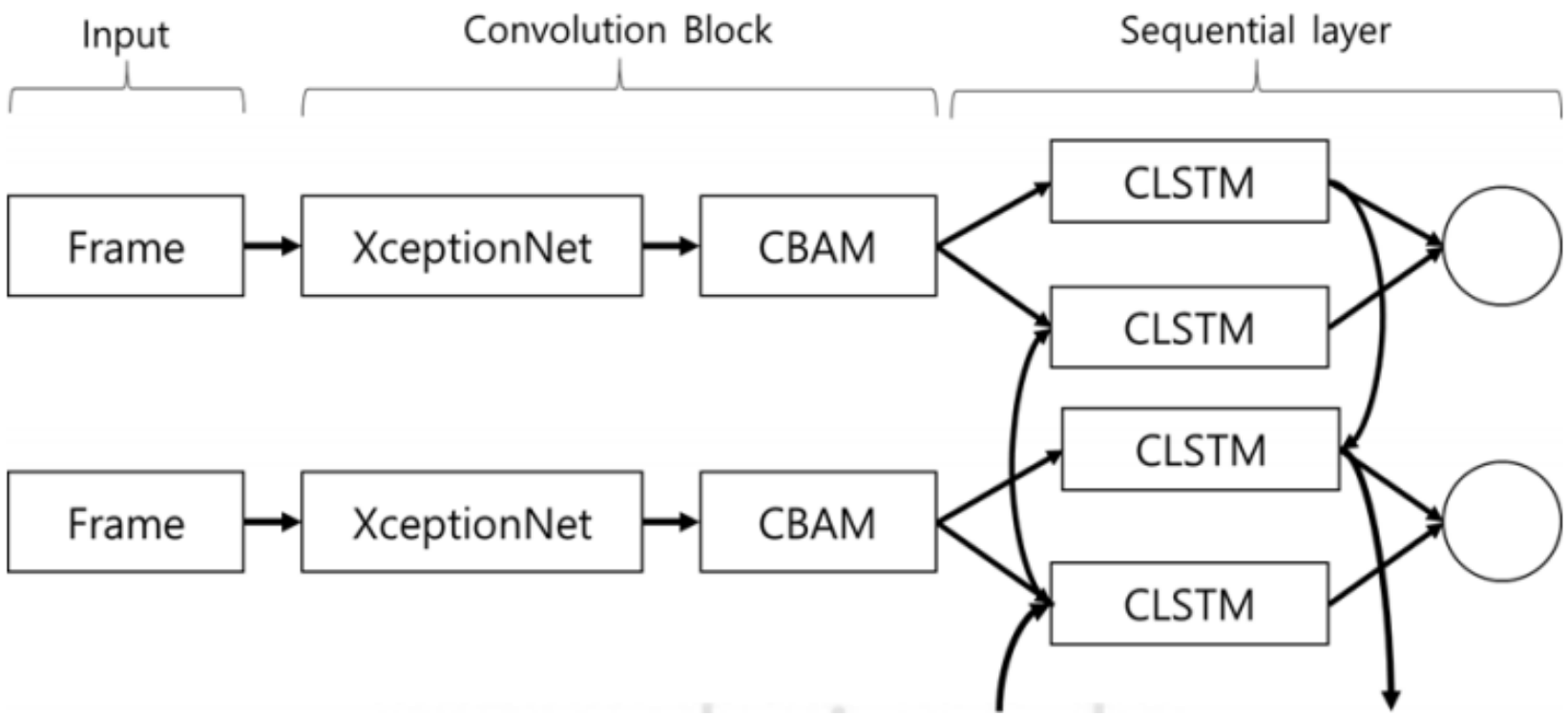
Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법

- Convolution Block에서는 프레임 단위로 분할된 동영상의 이미지 데이터로부터 각 프레임의 특징을 추출한다. - 동영상의 프레임별 특징을 추출하기 위해 합성곱 신경망 모델 중 하나인 **XceptionNet**과 중요한 특징에 가중치를 주어 학습하는 **어텐션 모듈(CBAM)**을 이용한다.



- 먼저 프레임 간의 연결 관계를 학습시키기 위해 XceptionNet을 사용하여 각 프레임의 특징을 추출하고, 추출된 프레임의 특징을 Convolutional Block Attention Module(CBAM)에 전달하여 CBAM을 학습시킨다. CBAM은 어텐션 모듈 중의 하나로써 가짜 동영상 판별에 큰 영향을 미치는 특징에 가중치를 주어 모델의 정확도를 높인다.

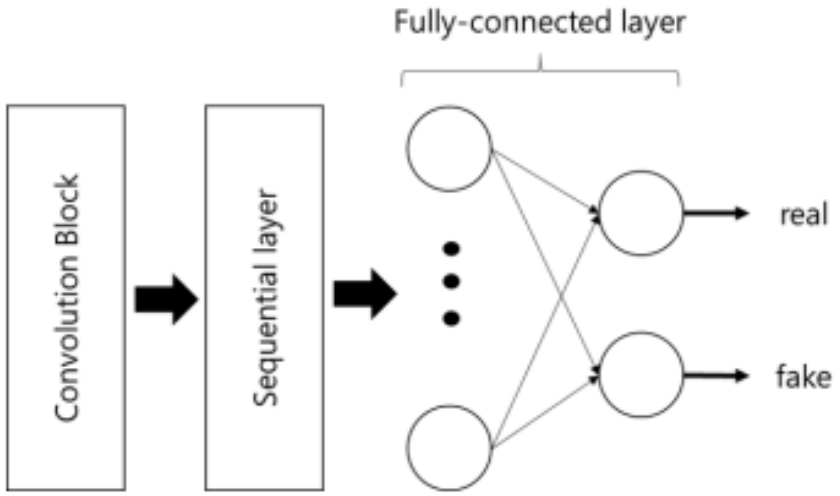
Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법



- Sequential layer는 **Bidirectional Convolutional LSTM을 사용**

– 인접한 프레임에 대해서 **프레임 간 연결이 자연스럽지 않고 일관적이지 않은 특징을 학습**. 또한 인접한 프레임의 불연속적인 특징을 학습할 때, **시간의 순방향 뿐만 아니라 역방향도 학습**

Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법



- Fully-connected layer에서는 Sequential layer의 출력을 입력 받아 Softmax 함수를 사용하여 계산된 값 중 큰 값을 이용하여 **실제 또는 가짜 동영상인지 판정**한다.

Bidirectional Convolutional LSTM을 이용한 Deepfake 탐지 방법

- 실험구성 : 실험은 Celeb-DF 데이터 셋을 학습하여 딥페이크 동영상 탐지를 수행하는 방식으로 진행
- 실험결과 : 본 논문에서 제안한 모델의 정확도는 93.5%으로 기존 제안되었던 모델들의 정확도보다 최대 30% 높은 것을 확인

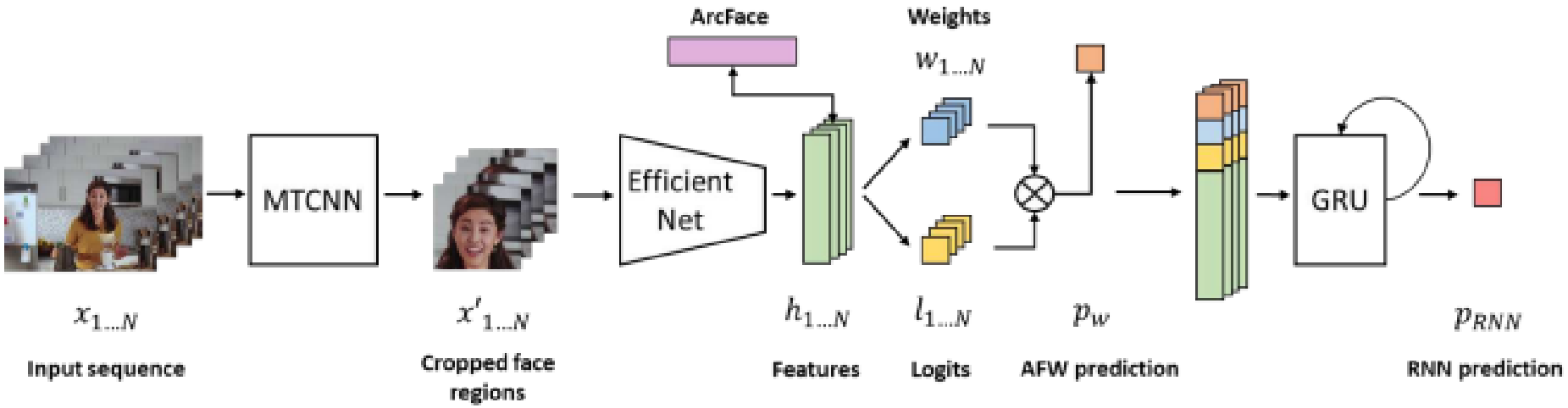
Method	Precision	Recall	F1-score	Acc.
our method	98.9%	88.0%	93.1%	93.5%

- 본 논문에서는 인접한 프레임의 불연속적인 특징을 **시간의 순방향 뿐만 아니라 역방향으로도 학습하는 딥페이크 탐지 모델을 제안**하였다. 기존 합성곱 신경망과 LSTM 셀을 단순하게 연결한 모델보다 Convolutional LSTM을 이용하면 딥페이크 탐지 시에 높은 정확도를 나타내는 것을 확인.

Deepfakes Detection with Automatic Face Weighting

- 본 논문에서는 영상에서 얼굴 조작을 정확하게 감지하기 위해 **CNN(Convolutional Neural Network)과 RNN(Recurrent Neural Network)을 결합한 새로운 모델 아키텍처를 제시**한다.
- 네트워크는 동영상이 실제 또는 가짜일 최종 확률을 제공하는 게이트 순환 장치(GRU)와 결합된 가중 메커니즘으로 이러한 조작을 탐지하기 위해 가장 신뢰할 수 있는 프레임을 자동으로 선택한다.

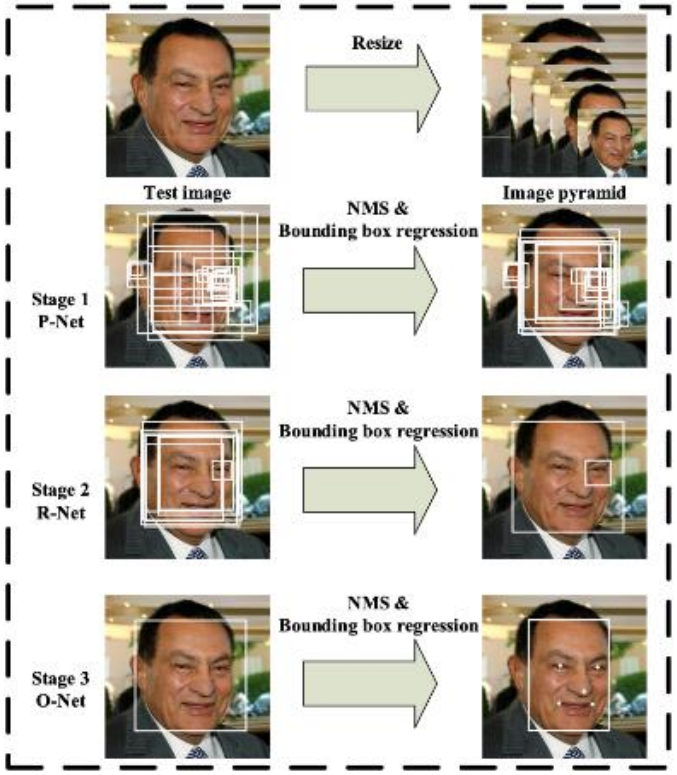
Deepfakes Detection with Automatic Face Weighting



- 이 방법에는 3가지 단계가 존재 → **(1) MTCNN을 사용한 여러 프레임의 얼굴 감지, (2) CNN을 사용한 특징 추출, (3) 게이트 순환 장치(GRU)와 함께 자동 얼굴 가중치(AFW)라고 하는 레이어를 사용한 예측 추정**

Deepfakes Detection with Automatic Face Weighting

- 먼저, MTCNN을 사용하여 얼굴 감지를 수행



- MTCNN은 그림과 같이 동작을 진행합니다. 총 3개의 스테이지로 구성되어 있고, 각각이 **P-Net, R-Net, O-Net**을 의미합니다.
- CNN의 앞의 MT는 멀티-태스크를 말하며 얼굴 검출, 랜드마크 검출, 그리고 상단의 사각형 모양을 말하는 Bounding box regression 세 개의 태스크를 함께 학습하는 Joint learning 방식을 사용하고 있습니다.

Deepfakes Detection with Automatic Face Weighting

- 얼굴 영역을 감지한 후, 실제 또는 가짜 얼굴을 분류하는 데 사용할 수 있는 특징을 추출하도록 이진 분류 모델을 훈련
- EfficientNet-b5를 사용하고 **EfficientNet을 일반적인 softmax + cross entropy loss 대신 ArcFace와 결합**
- ArcFace 손실은 마진, 즉 정규화된 가중치 및 기능 덕분에 하이퍼스피어의 측지 공간에서 결정 경계를 최대화하는데 이것은 얼굴 인식을 위한 매우 구별되는 기능을 얻고 무시할 수 있는 계산 오버헤드로 쉽게 구현할 수 있음.

Deepfakes Detection with Automatic Face Weighting

- 비디오 수준 예측을 결정할 때 얼굴이 감지된 가장 신뢰할 수 있는 영역을 강조하고 가장 신뢰할 수 없는 영역을 버리는 **자동 가중치 메커니즘을 사용하여 예측 추정**
- 모든 얼굴 영역과 프레임의 특징을 병합하기 위해 자동 얼굴 가중치 위에 순환 신경망(RNN)을 포함한다. 게이트 순환 장치(GRU)를 사용하여 모든 면 영역의 특징, 로짓 및 가중치를 결합하여 최종 추정치를 얻는다.
- GRU의 마지막 레이어의 출력은 선형 레이어와 Sigmoid 함수를 통해 매핑되어 조작되는 비디오의 최종 확률을 추정한다.

Deepfakes Detection with Automatic Face Weighting

- 실험구성 : DFDC 데이터 세트를 사용하여 방법을 교육하고 평가. 또한 제시된 접근 방식을 다른 4가지 기술과 비교

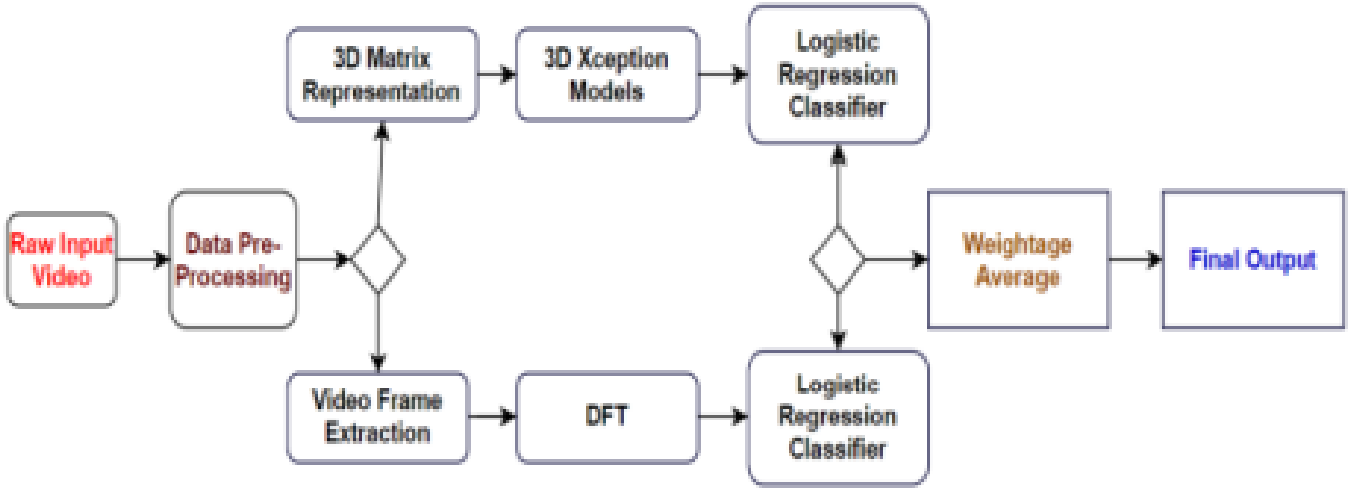
Method	Validation	Test
Conv-LSTM [30]	55.82%	57.63%
Conv-LSTM [30] + MTCNN	66.05%	70.78%
EfficientNet-b5 [42]	79.25%	80.62%
Xception [37]	78.42%	80.14%
Ours	92.61%	91.88%

- 실험결과 : 본 논문의 연구가 자동 얼굴 가중 계층과 GRU를 포함함으로써 정확도가 더욱 향상된다는 것을 보여준다.

DeepFake Detection using 3D-Xception Net with Discrete Fourier Transformation

- 이산 푸리에 변환과 함께 3D 팽창 Xception Net을 사용하여 딥 페이크 비디오를 탐지하는 새로운 방법을 제안.
- 모델은 두 개의 병렬 스트림으로 구성되어 있습니다. **하나는 Xception 모델**로 구성되고 **다른 하나는 푸리에 변환 기반 분류기**로 구성

DeepFake Detection using 3D-Xception Net with Discrete Fourier Transformation



- Xception 스트림에서는 CNN 내부의 2차원 레이어를 사용하는 대신 이미 존재하는 높이와 너비 치수에 비디오 깊이에 대한 차원을 추가해 3차원 컨볼루션 레이어로 교체하여 비디오에서 기능의 일부를 추출하여 분류를 위한 입력 매개 변수로 사용하는 대신 **전체 비디오를 입력으로 가져가는 3D 컨볼루션 신경망 모델을 사용**

DeepFake Detection using 3D-Xception Net with Discrete Fourier Transformation

- 두 번째 스트림은 **비디오에서 병렬 주파수 영역을 분석하기 위해 도입된 이산 푸리에 변환 모듈로 구성**
  → 영상의 주파수 분석에 별도의 스트림을 사용하는 이유는 가장 인위적으로 생성된 동영상이 동영상이나 이미지의 공간적 특성을 복제할 수 있고, 비디오의 복제 빈도와 진폭 분포는 비디오가 인위적으로 조작되었는지 여부를 결정하는 데 활용할 수 있다.

DeepFake Detection using 3D-Xception Net with Discrete Fourier Transformation

- 두 스트림에 의해 출력이 주어지면, 최종 결과를 얻기 위해 두 스트림의 출력이 결합되고, 최종 확률은 두 스트림의 개별 확률의 가중 평균을 취하여 사전 결정된 임계 값 확률 값과 비교하여 비디오가 가짜인지 실제인지를 결정합니다.

DeepFake Detection using 3D-Xception Net with Discrete Fourier Transformation

- 실험 구성 : Celeb-DF라는 공개 벤치마크 데이터 세트로 제안된 모델의 성능을 테스트, 모델의 성능은 수신기 작동 특성 그래프의 AUC(Area Under Curve) 점수를 사용하여 측정
- 실험 결과 : 3D Xception 모듈과 주파수 영역 기반 푸리에 변환 모델을 결합하는 파이프라인이 2차원 알고리즘 보다 훨씬 더 우수하다는 것

Method	Dimension	CALEB-DF Training	ROC-AUC %
Two Stream	2D	NO	53.8
MESO4	2D	NO	54.8
MESOInception4	2D	NO	53.6
HEADPose	2D	NO	54.6
FWA	2D	NO	56.9
VA-MLP	2D	NO	55
VALogReg	2D	NO	55.1
Xception-raw	2D	NO	48.2
Xception-c23	2D	NO	65.3
Xception-c40	2D	NO	65.5
Mutli task	2D	NO	54.3
Capsule	2D	NO	57.5
DSP-FWA	2D	NO	64.6
DFT	3D	YES	66.8
X ception-Metric-Learning	3D	YES	99.2
RCN	3D	YES	74.87
R2Plus1D	3D	YES	99.43
I3D	3D	YES	97.59
MC3	3D	YES	99.3
R3D	3D	YES	99.73
3D Xception-DFT (The Proposed)	**3D**	**YES**	**98.81**

A lightweight 3D convolutional neural network for deepfake detection

- 게재된 저널 : INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS (03 June 2021)
- Abstract 정보 :

DeepFake 기술의 급속한 발전은 비디오 콘텐츠의 진위성에 큰 도전을 불러왔습니다. DeepFake 검출 방법 개발이 매우 중요한데, 그중에서도 3차원(3D) 컨볼루션 신경망(CNN)이 폭넓은 관심을 끌고 만족스러운 성능을 달성했습니다. 그러나 DeepFake 탐지를 위해 설계된 3D CNN은 거의 없고 매개 변수가 커서 메모리 및 스토리지 소모가 많습니다. **본 논문에서는 딥페이크 감지를 위해 경량 3D CNN을 제안**합니다. 채널 변환 모듈은 상위 레벨에서 훨씬 적은 파라미터로 특징을 추출하도록 설계되었습니다. 공간-시간 모듈 역할을 하는 3D CNN은 시간 차원의 공간 특징을 융합하기 위해 채택됩니다. 프레임 콘텐츠를 억제하고 프레임 텍스처를 강조하기 위해 입력 프레임에서 공간이 풍부한 모델 피쳐를 추출하여 공간-시간 모듈이 더 나은 성능을 달성할 수 있도록 지원합니다. 실험 결과에 따르면 제안된 네트워크의 매개 변수 수는 다른 네트워크의 매개 변수보다 훨씬 적으며 제안된 네트워크는 주류 딥페이크 데이터 세트에서 다른 최첨단 딥페이크 탐지 방법을 능가합니다.

Deepfake and Security of Video Conferences

- 게재된 저널 : International Conference on Computer Science and Engineering (UBMK) (15-17 Sept. 2021)
- Abstract 정보 :

딥러닝은 인터넷에서 인공적인 콘텐츠를 만드는 데 널리 사용됩니다. 마찬가지로 가짜 콘텐츠를 탐지하는 데도 사용됩니다. 딥러닝 알고리즘과 통합되어 만들어진 가짜 프레임을 딥페이크라고 합니다. 최근에는 악성 유저들이 딥페이크를 사용하여 정품 콘텐츠를 조작하여 다양한 공격을 수행하는 경향이 있습니다. 화상 회의 애플리케이션은 COvid-19 대유행 초기부터 온라인 화상 회의에서 딥페이크 모델을 사용하여 가짜 가상 ID를 생성하는 악의적인 사용자의 주요 타겟이었습니다. 우리는 **가짜 얼굴을 탐지하기 위해 화상 회의 어플리케이션과 통합될 수 있는 경량 딥페이크 탐지 모델을 제안**합니다. 실험 분석에 따르면 제안된 모델은 화상 회의에서 가짜 이미지를 탐지할 수 있는 허용 가능한 정확도를 제공합니다.

Fake-buster: a lightweight solution for deepfake detection

- 게재된 저널 : (1 August 2021)
- Abstract 정보 :

최근 비디오 조작 기술의 발전으로 합성 미디어 제작이 그 어느 때보다도 용이해졌습니다. 요즘 비디오 에디션은 너무 사실적이어서 미디어 콘텐츠의 진실성을 평가하기 위해 감각에만 의존할 수는 없습니다. 조작 동영상의 양이 6개월마다 두 배씩 증가하는 상황에서 인터넷 곳곳에서 공유되는 방대한 양의 미디어를 처리하고, 관련 동영상을 최대한 빨리 제거할 수 있는 정교한 도구가 필요해 허위 정보를 부추기거나 주류 미디어에 대한 신뢰를 떨어뜨리는 등의 잠재적 피해를 줄여야 합니다. 본 논문에서, 우리는 현대의 얼굴 조작 기술을 대상으로 하는 비디오 시퀀스의 얼굴 조작 감지 문제를 다룹니다. 우리의 방법에는 두 개의 네트워크, (1) 비디오에 포함된 얼굴을 추출하는 네트워크, (2) 얼굴이 조작되었음을 나타내는 얼굴 및 주변 컨텍스트를 고려하여 얼굴이 조작되었음을 나타내는 조작 인식 네트워크가 포함됩니다. 특히, 우리는 **가지치기 및 지식 증류와 같은 신경망 압축 기술을 사용하여 비디오 스트림을 신속하게 처리할 수 있는 경량 솔루션을 만들 것을 제안**합니다. 우리의 접근 방식은 인터넷에서 발견되는 유기적 콘텐츠를 반영하고 최첨단 딥페이크 탐지 접근 방식과 비교하여 5가지 다른 조작 기법으로 구성된 비디오로 구성된 DeepFake Detection Dataset에서 검증되었습니다.

Extracting Deep Local Features to Detect Manipulated Images of Human Faces

- 게재된 저널 : IEEE International Conference on Image Processing (ICIP) (2020)
- Abstract 정보 :

컴퓨터 비전과 머신러닝의 최근 발전은 인간의 얼굴을 사실적으로 조작한 비디오를 만드는 것을 가능하게 했고, 그러한 기능에 의해 풀리는 악의적 효과에 대한 적절한 보호를 보장하는 문제를 제기했습니다. **본 논문에서 우리는 조작된 얼굴 이미지의 자동 탐지를 위한 핵심 요소로 조작된 영역 간에 공유되는 로컬 이미지 특징을 제안**합니다. 또한 **이러한 특징을 추출하기 위한 올바른 구조적 편향을 가진 경량 아키텍처를 설계하고 이미지 클래스 감독만을 일관되게 능가하는 다중 작업 훈련 체계를 도출**합니다. 훈련된 네트워크는 크게 줄어든 매개변수를 사용하여 FaceForensics++ 데이터 세트에서 최첨단 결과를 달성하며 완전히 생성된 얼굴 이미지를 탐지하는 데 잘 작동하는 것으로 나타났습니다.