

(43) 공개일자 2024년07월04일

- 전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **셀프 어텐션 및 다중 작업 학습을 통한 주파수 기반 이미지 위조 위치 검출 방법 및 시스템**

일 실시예에 따른 위조 위치 검출 시스템에 의해 수행되는 이미지 위조 위치 검출 방법은, 주파수 기반 위조 검출 모델에 입력 이미지를 입력받는 단계; 및 상기 주파수 기반 위조 검출 모델을 통해 상기 입력받은 입력 이미지와 상기 입력받은 입력 이미지로부터 주파수 분해를 통해 획득된 저주파 이미지 및 고주파 이미지를 이용하여 상기 입력받은 입력 이미지의 위조 영역을 검출하는 단계를 포함할 수 있다.

Figure 1 illustrates the architecture of the proposed method. The process begins with DCT Decomposition (110), which takes an input image and decomposes it into low, full, and high frequency components. These components are then processed by three parallel streams (e₁, e₂, e₃). Each stream consists of a Receptive Block (RB) and a feature fusion block (F). The outputs of these streams are concatenated (C) and passed through a Bridge and DFSAM blocks. A Region Mask and Label Mask are used to calculate L_{BCE} + L_{DICE} and L_{FL} + L_{DICE} losses, which are supervised (S) by a pre-trained segmentation model. The final output is a segmentation map.

Legend:

- $\textcircled{S}$: Dropout ($p=0.5$)
- $\textcircled{F}$: FAM
- $\textcircled{C}$: Concatenation
- $\textcircled{S}$: Pre-trained segmentation model
- $\dashrightarrow$: Supervision

(52) CPC특허분류

G06N 3/048 (2023.01)

G06N 3/082 (2023.01)

G06V 10/88 (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711162433
과제번호	2021R1A2C2010893
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	중견연구
연구과제명	[Ezbaro] 희소 영상에 특화된 영상 분해-합성을 통한 멀티도메인 기반 관심 지표 및
객체 검출 기술 연구	
기 여 율	1/2
과제수행기관명	인하대학교
연구기간	2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711174118
과제번호	00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간] 인공지능융합혁신인재양성(1차년도)
기 여 율	1/2
과제수행기관명	인하대학교 산학협력단
연구기간	2022.07.01 ~ 2022.12.31

명세서

청구범위

청구항 1

위조 위치 검출 시스템에 의해 수행되는 이미지 위조 위치 검출 방법에 있어서,

주파수 기반 위조 검출 모델에 입력 이미지를 입력받는 단계; 및

상기 주파수 기반 위조 검출 모델을 통해 상기 입력받은 입력 이미지와 상기 입력받은 입력 이미지로부터 주파수 분해를 통해 획득된 저주파 이미지 및 고주파 이미지를 이용하여 상기 입력받은 입력 이미지의 위조 영역을 검출하는 단계

를 포함하는 이미지 위조 위치 검출 방법.

청구항 2

제1항에 있어서,

상기 주파수 기반 위조 검출 모델은,

종단 간(end-to-den) 학습을 위한 인코더 네트워크 및 디코더 네트워크를 포함하며, 확장 주파수 셀프 어텐션 모듈(Dilated Frequency Self-Attention Module; DFSAM)로 마스크 영역과 마스크 엣지를 동시에 학습하도록 구성된 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 3

제2항에 있어서,

상기 주파수 기반 위조 검출 모델은,

입력 이미지에 대한 DCT 변환 함수를 이용하여 DCT 변환 이미지를 생성하고, 상기 생성된 DCT 변환 이미지에서 고주파 성분을 추출하기 위해 저주파 영역을 필터링하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 4

제2항에 있어서,

상기 인코더 네트워크는,

각 인코더 블록에 특정 크기의 커널, 스트라이드, 패딩의 컨볼루션 레이어, 배치 정규화 레이어, ReLU 활성화 함수로 구성되고,

상기 디코더 네트워크는,

각 디코더 블록에 특정 크기의 커널, 스트라이드, 패딩, ReLU 활성화 함수, 인접 업샘플링 레이어, 특정 값 확률의 드롭아웃으로 구성되어 다중작업 학습을 수행하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 5

제3항에 있어서,

상기 주파수 기반 위조 검출 모델은,

입력 이미지로부터 주파수 분해를 통해 저주파 이미지 및 고주파 이미지를 획득하고, 상기 획득된 저주파 이미지와 고주파 이미지가 공유 인코더를 통해 서로 다른 주파수 유형의 이미지를 학습하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 6

제5항에 있어서,

상기 주파수 기반 위조 검출 모델은,

상기 학습을 통해 획득된 인코딩 정보를 어텐션 기반의 브릿지 레이어로 전달하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 7

제2항에 있어서,

상기 확장 주파수 셀프 어텐션 모듈은,

셀프 어텐션 맵을 생성하기 위해 상기 인코더 블록을 통해 출력된 3개의 출력 특징맵을 연결하고, 특정 크기의 커널인 컨볼루션 레이어를 적용하여 채널의 개수를 감소시키는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 8

제6항에 있어서,

상기 확장 주파수 셀프 어텐션 모듈은,

다른 출력 특징맵을 특정 크기의 커널인 개별 단일 컨볼루션 레이어에 공급하여 새로운 특징들을 생성하고, 상기 생성된 새로운 특징들 중 제1 새로운 특징맵과 제2 새로운 특징맵 사이의 전치 사이에 행렬 곱셈을 수행한 후 소프트웨어 연산을 수행하여 셀프 어텐션 맵을 획득하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 9

제2항에 있어서,

상기 주파수 기반 위조 검출 모델은,

디코더 블록과 대응되는 인코더 블록으로부터 추출된 동일한 공간 크기의 특징맵에서 디코더 블록으로 스킵 연결을 수행하되, 상기 특징맵을 인코더 블록에서 디코더 블록으로 스킵 연결하기 전에 잔차 블록(residual block; RB)을 적용하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 10

제9항에 있어서,

상기 주파수 기반 위조 검출 모델은,

사전 훈련 분할 모델에서 입력 이미지에서 분할된 배경과 객체에 대한 정보를 포함하는 보조 특징을 통해 밀집 영역의 분할 결과를 획득하고, 특징 어텐션 모듈(Feature Attention Module; FAM)에서 상기 특징맵에 대해 채널 간 특징 관계를 제공하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 11

제10항에 있어서,

상기 주파수 기반 위조 검출 모델은,

레이블의 영역 및 엣지 정보의 다중작업 학습을 통해 위조된 영역의 위치가 검출되도록 영역 예측과 엣지 예측을 출력하는 것을 특징으로 하는 이미지 위조 위치 검출 방법.

청구항 12

위조 위치 검출 시스템에 있어서,

주파수 기반 위조 검출 모델에 입력 이미지를 입력받는 이미지 입력부; 및

상기 주파수 기반 위조 검출 모델을 통해 상기 입력받은 입력 이미지와 상기 입력받은 입력 이미지로부터 주파수 분해를 통해 획득된 저주파 이미지 및 고주파 이미지를 이용하여 상기 입력받은 입력 이미지의 위조 영역을 검출하는 위조 영역 검출부

를 포함하는 이미지 위조 위치 검출 시스템.

발명의 설명

기술 분야

[0001] 아래의 설명은 이미지 위조 검출 기술에 관한 것이다.

배경 기술

[0003] 지난 몇 년 동안 그래픽 편집 기술의 급속한 발전으로 공격자는 멀티미디어를 조작, 위조, 변경하여 인간의 시각 시스템에 보이지 않게 만들었다. 위조된 이미지들은 더 유행하고 정교해졌다. 특히 스플라이싱과 복사 이동은 가장 널리 사용되는 기술로, 사람의 눈에 거의 감지할 수 없는 방식으로 이미지를 처리한다. 한편, 이러한 방식으로 만들어진 가짜 미디어의 악의적인 확산은 가짜 뉴스 전파, 거래 사기, 개인의 사생활 침해 등 우리 사회에 심각한 피해를 입힌다. 따라서, 위조된 디지털 이미지의 악의적인 전파를 방지하기 위한 효과적인 이미지 위조 감지 방법은 현대 사회에서 중요한 문제로 대두되고 있다.

[0004] 한편, 대한민국 공개특허 제10-2021-0147370호(2021.12.07 공개)에 위조된 모조 이미지에 대하여 변조 및 위조 여부를 탐지할 수 있는 이미지 위조 탐지 장치 및 방법이 개시되어 있다.

발명의 내용

해결하려는 과제

[0006] 종단 간 학습이 가능한 인코더-디코더 구조로 구성되어 셀프 어텐션 및 다중작업 학습을 수행하는 주파수 기반 위조 검출 모델을 이용하여 이미지의 위조 영역을 검출하는 방법 및 시스템을 제공할 수 있다.

과제의 해결 수단

[0008] 위조 위치 검출 시스템에 의해 수행되는 이미지 위조 위치 검출 방법은, 주파수 기반 위조 검출 모델에 입력 이미지를 입력받는 단계; 및 상기 주파수 기반 위조 검출 모델을 통해 상기 입력받은 입력 이미지와 상기 입력받은 입력 이미지로부터 주파수 분해를 통해 획득된 저주파 이미지 및 고주파 이미지를 이용하여 상기 입력받은 입력 이미지의 위조 영역을 검출하는 단계를 포함할 수 있다.

[0009] 상기 주파수 기반 위조 검출 모델은, 종단 간(end-to-den) 학습을 위한 인코더 네트워크 및 디코더 네트워크를 포함하며, 확장 주파수 셀프 어텐션 모듈(Dilated Frequency Self-Attention Module; DFSAM)로 마스크 영역과 마스크 옛지를 동시에 학습하도록 구성된 것일 수 있다.

[0010] 상기 주파수 기반 위조 검출 모델은, 입력 이미지에 대한 DCT 변환 함수를 이용하여 DCT 변환 이미지를 생성하고, 상기 생성된 DCT 변환 이미지에서 고주파 성분을 추출하기 위해 저주파 영역을 필터링하는 것일 수 있다.

[0011] 상기 인코더 네트워크는, 각 인코더 블록에 특정 크기의 커널, 스트라이드, 패딩의 컨볼루션 레이어, 배치 정규화 레이어, ReLU 활성화 함수로 구성되고, 상기 디코더 네트워크는, 각 디코더 블록에 특정 크기의 커널, 스트라이드, 패딩, ReLU 활성화 함수, 인접 업샘플링 레이어, 특정 값 확률의 드롭아웃으로 구성되어 다중작업 학습을 수행하는 것일 수 있다.

[0012] 상기 주파수 기반 위조 검출 모델은, 입력 이미지로부터 주파수 분해를 통해 저주파 이미지 및 고주파 이미지를 획득하고, 상기 획득된 저주파 이미지와 고주파 이미지가 공유 인코더를 통해 서로 다른 주파수 유형의 이미지를 학습하는 것일 수 있다.

[0013] 상기 주파수 기반 위조 검출 모델은, 상기 학습을 통해 획득된 인코딩 정보를 어텐션 기반의 브릿지 레이어로 전달하는 것일 수 있다.

- [0014] 상기 확장 주파수 셀프 어텐션 모듈은, 셀프 어텐션 맵을 생성하기 위해 상기 인코더 블록을 통해 출력된 3개의 출력 특징맵을 연결하고, 특정 크기의 커널인 컨볼루션 레이어를 적용하여 채널의 개수를 감소시키는 것일 수 있다.
- [0015] 상기 확장 주파수 셀프 어텐션 모듈은, 다른 출력 특징맵을 특정 크기의 커널인 개별 단일 컨볼루션 레이어에 공급하여 새로운 특징들을 생성하고, 상기 생성된 새로운 특징들 중 제1 새로운 특징맵과 제2 새로운 특징맵 사이의 전치 사이에 행렬 곱셈을 수행한 후 소프트맥스 연산을 수행하여 셀프 어텐션 맵을 획득하는 것일 수 있다.
- [0016] 상기 주파수 기반 위조 검출 모델은, 디코더 블록과 대응되는 인코더 블록으로부터 추출된 동일한 공간 크기의 특징맵에서 디코더 블록으로 스킵 연결을 수행하되, 상기 특징맵을 인코더 블록에서 디코더 블록으로 스킵 연결하기 전에 잔차 블록(residual block; RB)을 적용하는 것일 수 있다.
- [0017] 상기 주파수 기반 위조 검출 모델은, 사전 훈련 분할 모델에서 입력 이미지에서 분할된 배경과 객체에 대한 정보를 포함하는 보조 특징을 통해 밀집 영역의 분할 결과를 획득하고, 특징 어텐션 모듈(Feature Attention Module; FAM)에서 상기 특징맵에 대해 채널간 특징 관계를 제공하는 것일 수 있다.
- [0018] 상기 주파수 기반 위조 검출 모델은, 레이블의 영역 및 엣지 정보의 다중작업 학습을 통해 위조된 영역의 위치가 검출되도록 영역 예측과 엣지 예측을 출력하는 것일 수 있다.
- [0019] 위조 위치 검출 시스템은, 주파수 기반 위조 검출 모델에 입력 이미지를 입력받는 이미지 입력부; 및 상기 주파수 기반 위조 검출 모델을 통해 상기 입력받은 입력 이미지와 상기 입력받은 입력 이미지로부터 주파수 분해를 통해 획득된 저주파 이미지 및 고주파 이미지를 이용하여 상기 입력받은 입력 이미지의 위조 영역을 검출하는 위조 영역 검출부를 포함할 수 있다.

발명의 효과

- [0021] 셀프 어텐션 및 다중 작업 학습을 위한 주파수 기반 위조 위치 검출 모델을 통해 엣지와 영역을 동시에 예측하여 위조 위치 검출의 정확도를 향상시킬 수 있다.

도면의 간단한 설명

- [0023] 도 1은 일 실시예에 있어서, 주파수 기반 위조 검출 모델을 설명하기 위한 도면이다.
- 도 2는 일 실시예에 있어서, 확장 주파수 셀프 어텐션 모듈을 설명하기 위한 도면이다.
- 도 3은 일 실시예에 있어서, 매개변수의 변화에 따라 이진 마스크 생성을 시각화한 도면이다.
- 도 4는 일 실시예에 있어서, 위조 위치 검출 시스템의 구성을 설명하기 위한 블록도이다.
- 도 5는 일 실시예에 있어서, 위조 위치 검출 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.
- [0026] 도 1은 일 실시예에 있어서, 주파수 기반 위조 검출 모델을 설명하기 위한 도면이다.
- [0027] 위조 위치 검출 시스템은 주파수 기반 위조 검출 모델을 이용하여 입력 이미지에 대한 위조 영역의 위치를 검출할 수 있다. 주파수 기반 위조 검출 모델은 중단 간 학습 아키텍처를 갖춘 컨볼루션 인코더-디코더 네트워크로 구성될 수 있다. 주파수 기반 위조 검출 모델은 확장 주파수 셀프 어텐션 모듈(Dilated Frequency Self-Attention Module; DFSAM)로 마스크 영역과 마스크 엣지를 동시에 학습하는 새로운 멀티 학습 네트워크를 제공할 수 있다.
- [0028] 주파수 기반 위조 검출 모델은 DCT(Discrete Cosine Transformation) 필터링을 통해 주파수 정보를 추출하며, 주파수 셀프 어텐션 모듈을 통해 주파수 정보를 결합하여 위조된 영역의 흔적을 강조할 수 있으며, 디코딩을 수

행하는 동안 특정 위조 객체의 위치를 안내하기 위해 사전 훈련된 분할 모델을 통합한다. 이에, 엣지와 영역을 동시에 예측하여 위조 위치 검출의 정확도를 향상시킬 수 있다.

주파수 기반 위조 검출 모델의 표기법과 알고리즘은 표 1과 알고리즘 1을 참고하기로 한다.

표 1:

Notation	Description
$\mathbf{x}^{full}$	input RGB domain image
$\mathbf{x}^{low}, \mathbf{x}^{high}$	low-/high- frequency decomposed image
$\mathcal{D}$	Discrete Cosine Transformation function
$\mathcal{D}^{-1}$	Inverse Discrete Cosine Transformation function
$\mathbf{x}^d$	DCT transformed image
$\mathbf{I}_\alpha^{low}, \mathbf{I}_\alpha^{high}$	indicator function for mask generation
α	hyperparameter for mask generation
$\otimes$	element-wise multiplication
e_i	encoder blocks, $i = 1, 2, 3$
$\mathbf{x}_i^{full}, \mathbf{x}_i^{low}, \mathbf{x}_i^{high}$	output of each encoder block, $i = 1, 2, 3$
$\mathbf{X}$	concatenated features of last encoder output
$\mathbf{Q}, \mathbf{K}, \mathbf{V}$	self-attention features
$\mathbf{A}$	self-attention map
$\odot$	matrix multiplication
$\mathbf{O}$	output of DFSAM
y	2D dilated convolutional filter
r	dilation rate
d_i	decoder blocks, $i = 1, 2, 3$
$\mathbf{P}_i$	output of decoder block, $i = 1, 2, 3$
$\mathbf{S}$	auxiliary segmentation model
s_i	auxiliary segmentation feature
$\hat{\mathbf{R}}$	region prediction
$\hat{\mathbf{E}}$	edge prediction

알고리즘 1:

Algorithm 1 FBI-Net Forward Procedure.

Require: Forged image $\mathbf{x}^{full}$, Low-frequency range parameter α

Apply DCT on forged image, $\mathbf{x}^d \leftarrow \mathcal{D}(\mathbf{x}^{full})$

Define filter masks, $\mathbf{I}_\alpha^{low}$, and $\mathbf{I}_\alpha^{high}$

$\mathbf{x}^{low}, \mathbf{x}^{high} \leftarrow \mathcal{D}^{-1}(\mathbf{I}_\alpha^{low} \otimes \hat{\mathbf{x}}), \mathcal{D}^{-1}(\mathbf{I}_\alpha^{high} \otimes \hat{\mathbf{x}})$

for $i = 1, 2, 3$ **do**

for $z = low, high, full$ **do**

$\mathbf{x}_i^z \leftarrow e_i(\mathbf{x}_{i-1}^z)$ $\triangleright \mathbf{x}_0^z = \mathbf{x}^z$

$\mathbf{P}_0 \leftarrow \text{DFSAM}([\mathbf{x}_3^{low}, \mathbf{x}_3^{high}, \mathbf{x}_3^{full}])$

for $i = 1, 2, 3$ **do**

$s_i \leftarrow \text{Conv}(\mathbf{S}(\mathbf{x}^{full}))$

$\mathbf{P}_i \leftarrow d_i(\text{FAM}([RB(\mathbf{x}_{4-i}^{full}), \mathbf{P}_{i-1}, s_i]))$

$\hat{\mathbf{R}} \leftarrow C_{3 \times 3}(\mathbf{P}_3; w_{\hat{\mathbf{R}}})$

$\hat{\mathbf{E}} \leftarrow C_{3 \times 3}(\mathbf{P}_3; w_{\hat{\mathbf{E}}})$

return Region Prediction $\hat{\mathbf{R}}$, Edge Prediction $\hat{\mathbf{E}}$

주파수 기반 위조 검출 모델에 저주파/고주파 구성 요소를 이용하여 위조 및 진짜 이미지 영역을 분류하는 네트워크가 구성될 수 있다. 주파수 기반 위조 검출 모델은 입력 이미지로부터 저주파/고주파 구성 요소로 위조 및 진짜 이미지 영역을 분류할 수 있다. 입력 이미지의 주파수를 분해하기 위해 계산 복잡도가 $O(N^2 \log 2N)$ 인 2D

DCT(110) 변환 함수가 사용될 수 있다. $\mathbf{x}^{full} \in \mathbb{R}^{H \times W \times 3}$ 가 입력 이미지로 정의될 수 있으며, 여기서, H와 W가 입력 이미지의 $\mathbf{x}^{full}$ 의 높이와 너비를 나타낸다. 먼저, RGB 도메인 입력이 DCT 도메인으로 변환될 수 있다.

수학식 1:

$$\hat{\mathbf{x}} = \mathcal{D}(\mathbf{x}^{full}),$$

여기서, $\hat{\mathbf{x}} \in \mathbb{R}^{H \times W \times 3}$ 는 DCT 변환 이미지, D는 2D DCT 변환 함수를 나타낸다.

고주파 성분을 추출하기 위해 $\hat{\mathbf{x}}$ 의 저주파 영역이 필터링될 수 있다. 고주파 성분이 $\mathbf{x}^{full}$ 에 포함된 왜곡을 강조하는데 도움이 될 수 있다. 또한, 저주파 성분은 일반적인 성분에서 발견되는 대부분의 에너지를 유지하기 위해 필수적인 주파수 정보이다. 실시예에서는 고주파 성분과 저주파 성분을 포함하는 두 가지 유형의 정보를 사용하여 이미지 왜곡과 전체 콘텐츠를 활용할 수 있다. 단일 RGB 이미지를 입력 이미지로 사용할 때보다 고품질의 정보를 보다 효과적으로 학습할 수 있을 것으로 기대된다. 고주파 구성 요소와 저주파 구성 요소를 필터링하기 위해 지표 함수 $\mathbf{I}_{\alpha}^{low}$, $\mathbf{I}_{\alpha}^{high} \in \mathbb{R}^{H \times W}$ 가 다음과 같이 정의될 수 있다.

수학식 2:

$$\mathbf{I}_{\alpha}^{low}(i, j) = \begin{cases} 1 & \text{if } 0 \leq i \leq W - \alpha \text{ and} \\ & 0 \leq j \leq H - \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

수학식 3:

$$\mathbf{I}_{\alpha}^{high} = 1 - \mathbf{I}_{\alpha}^{low},$$

여기서, α 는 저주파의 범위를 결정하는 초매개변수를 나타낸다. 이진 마스크는 사각 행렬의 왼쪽 상단과 오른쪽 하단 영역을 구분하며, 여기서, α 값은 마스크의 왼쪽 상단 모서리에 있는 (0, 0) 축에서 시작하여 완전한 이등변 삼각형을 형성한다. 예를 들면, 로우패스 필터 마스크를 사용하면, α 값이 H와 W일 때, 오른쪽 하단 모서리 값이 0이고, 왼쪽 상단 모서리 값이 1인 이진 필터가 생성될 수 있다. 또한 하이패스 필터 마스크는 로우패스 필터 마스크에서 1을 뺀 것과 같다. 도 3을 참고하면, 하이패스 필터 마스크는 첫 번째 행에 표시되고, 로우패스 필터 마스크는 세 번째 행에 표시될 수 있다. 해당 필터링 이미지는 마지막 행에 표시될 수 있다.

$\mathbf{x}^d$ 는 이진 마스크 $\mathbf{I}_{\alpha}^{low}$ 와 $\mathbf{I}_{\alpha}^{high}$ 의 요소별 곱에 의해 분해된다. 다음과 같이 역 이산 코사인 변환(inverse Discrete Cosine Transform; IDCT)을 적용하여 $\mathbf{x}^{low}, \mathbf{x}^{high} \in \mathbb{R}^{H \times W \times 3}$ 이 획득될 수 있다.

수학식 4:

$$\mathbf{x}^{low} = \mathcal{D}^{-1}(\mathbf{I}_{\alpha}^{low} \otimes \hat{\mathbf{x}})$$

수학식 5:

$$\mathbf{x}^{high} = \mathcal{D}^{-1}(\mathbf{I}_{\alpha}^{high} \otimes \hat{\mathbf{x}})$$

여기서, $\otimes$ 와 $\mathcal{D}^{-1}$ 는 각각 요소별 곱과 역 이산 코사인 변환(IDCT)을 나타낸다.

주파수 기반 위조 검출 모델은 백본 네트워크로 ImageNet을 이용하여 사전 훈련된 ResNet18과 함께 인코더-디코

더 네트워크로 U-Net가 사용될 수 있다. 보다 조밀한 특징맵을 유지하고 효율적인 네트워크를 설계하기 위해 ResNet의 마지막 다운샘플링 연산이 제거될 수 있다. 이를 통해 출력 특징맵의 해상도를 입력 이미지의 1/8로 높일 수 있다.

각 인코더 블록 e_i 는 커널 크기 3×3 , 스트라이드 1, 패딩 1의 컨볼루션 레이어와 배치 정규화 레이어, ReLU 활성화 함수로 구성된다. 각 인코더 블록 e_i 는 3개의 입력 $\mathbf{x}^{low}, \mathbf{x}^{high}, \mathbf{x}^{full}$, 출력 $\mathbf{x}_i^{low}, \mathbf{x}_i^{high}, \mathbf{x}_i^{full}$ 과 함께 매개변수를 공유한다($i=1,2,3$).

수학식 6:

$$\mathbf{x}_i^z = e_i(\mathbf{x}_{i-1}^z)$$

여기서 $z=low, high, full$ 일 때, $\mathbf{x}_0^z = \mathbf{x}^z$ 이다.

단일 인코더에서 서로 다른 주파수 유형의 이미지를 학습함으로써 인코더는 다양한 입력 범위 측면에서 풍부한 표현을 학습할 수 있다.

도 2를 참고하면, 확장 주파수 셀프 어텐션 모듈(Dilated Frequency Self-Attention Module; DFSAM)(120)에 대하여 설명하기로 한다. 입력 이미지의 인코딩 단계에서 마지막 레이어는 추상적이고 고차원적인 정보가 포함된다. 인코딩된 정보는 어텐션 기반 브릿지 레이어로 전달된다. RGB 정보 $\mathbf{x}_3^{full}$ 는 가짜 이미지에서 비정상적인 텍스처를 식별하는 데 도움이 되며, 압축된 주파수 정보 $\mathbf{x}_3^{low}$ 및 $\mathbf{x}_3^{high}$ 는 미묘한 바이어스 오류를 개선한다. 실시예에서는 어텐션 맵을 추출하고 필수 정보를 강조하는 확장 주파수 셀프 어텐션 모듈이 제안될 수 있다.

확장 주파수 셀프 어텐션 모듈의 구조는 도 2와 같다. 셀프 어텐션 맵 A를 생성하기 위해 먼저 3개의 출력 특징맵 $\mathbf{X} = [\mathbf{x}_3^{low}, \mathbf{x}_3^{high}, \mathbf{x}_3^{full}] \in \mathbb{R}^{H' \times W' \times 3C'}$ 을 연결하고 커널 크기가 1×1 인 컨볼루션 레이어를 $\mathbf{X}' \in \mathbb{R}^{H' \times W' \times C'}$ 에 적용하여 채널을 줄인다. 그런 다음 $\mathbf{X}'$ 을 커널 크기가 3×3 인 3개의 개별 단일 컨볼루션 레이어에 공급하여 3개의 새로운 특징(Q, K, V)를 생성할 수 있다. 그런 다음, $\mathbb{R}^{C' \times N}$ 으로 재구성할 수 있다. 여기서 $N=H' \times W'$ 는 각 특징맵의 픽셀 수이다. Q와 K의 전치 사이에 $\odot$ 로 표시된 행렬 곱셈을 수행하고 소프트맥스 연산을 적용하여 셀프 어텐션 맵 $\mathbf{A} \in \mathbb{R}^{C' \times C'}$ 을 계산할 수 있다.

수학식 7:

$$A_{ij} = \frac{\exp(\max(Q_i \odot K_j))}{\sum_{i=1}^N \exp(\max(Q_i \odot K_j))}$$

여기서, A_{ij} 는 j번째 채널에 대한 i번째 채널의 영향을 측정한다. 어텐션 메커니즘을 생성하기 위해 A와 V 사이의 행렬 곱셈을 수행할 수 있다. 그런 다음, 출력 O_j 를 도출하기 위해 A와 스케일링된 특징맵 사이에 요소별 합을 적용한다.

수학식 8:

$$O_j = \sum_{i=1}^N (V_{ji} \odot A_i)$$

O_j 가 $\mathbf{O} \in \mathbb{R}^{H' \times W' \times C'}$ 로 변형될 수 있다.

다중 스케일 특징 융합은 필터의 필드를 넓히기 위해 서로 다른 팽창률을 갖는 컨볼루션 레이어를 적용하여 다

양한 스케일에서 위조된 영역을 감지한다. 위조된 이미지에서 다양한 크기의 객체를 발견했기 때문에 아트로스 특징맵(Atrous Feature Map)에 다중 스케일 아트로스 컨볼루션 퓨전(atrous convolution fusion)이 적용될 수 있다. 수학적으로 y 로 표시된 2D 확장 컨볼루션 필터는 다음과 같이 정의될 수 있다.

수학식 9:

$$y(i) = \sum_k x(i + rk)w(k)$$

여기서, x 와 w 는 각각 입력 신호와 필터이고, r 은 입력 신호의 샘플링 단계 크기를 결정하는 팽창률이다. 실시예에서는 팽창률 $r = [1, 6, 12, 18]$ 이라 설정하고, 각 팽창률에 대한 결과 특징맵을 연결할 수 있다. $r = 1$ 이면 표준 컨볼루션 연산과 동일하다.

다시 도 1을 참고하면, 다중작업 학습(Multi-Task Learning; MTL)을 수행하는 디코더는 수많은 관련 작업에서 표현을 공유함으로써 심층 네트워크의 학습 효율성을 향상시킬 수 있다. 디코더는 특징맵을 추가로 처리하여 최종 픽셀별 예측을 생성할 수 있다.

도 1에 표시된 디코더 블록 d_i 는 시퀀스가 하단에서 시작하여 $i = 1, 2, 3$ 에 대한 출력을 생성하는 마지막 디코더 블록으로 끝난다. 디코더 블록 d_i 는 커널 크기 3×3 , 스트라이드 1, 패딩 1, ReLU 활성화 함수, 비율 2의 가장 가까운 인접 업샘플링 레이어, 0.5 확률의 드롭아웃으로 구성될 수 있다. 그러나 이전 컨볼루션 블록의 특징맵이 업샘플링되면 컨볼루션 연산에 의해 추출된 위조 추적이 사라질 수 있다. 업샘플링으로 인한 특징 정보 손실을 보상하기 위해 디코더 블록이 있는 해당 단계 인코더 블록으로부터 추출된 동일한 공간 크기의 특징맵에서 디코더 블록으로 스킵 연결을 수행한다. 특징맵을 인코더 블록에서 디코더 블록으로 스킵 연결하기 전에 RB(residual block; 잔차 블록)이 적용될 수 있다. 잔차 블록(RB)은 배치 정규화 레이어를 제거하는 단순화된 사전 활성화 잔차 블록이다.

사전 훈련된 분할 모델은 디코더 훈련을 지원하도록 설계될 수 있다. 위조된 이미지에는 진짜 객체와 진짜 객체가 혼합되어 있다. 일반적인 데이터셋으로 훈련된 사전 훈련 분할 모델은 위조된 객체의 검출을 지원하는 보조 특징으로 모든 객체를 추출하는 경향이 있지만, 위조되지 않은 경우에도 진짜 객체를 위조 영역으로 감지하는 경향이 있다는 한계가 있다. 주파수 기반 위조 검출 모델은 획득한 객체 위치를 기반으로 진짜 객체를 식별하고 사전 훈련 분할 모델이 찾을 수 없는 진짜 배경을 확인한다. 21개 범주가 Resnet50을 백본 네트워크로 사용하여 DeepLabV3C+의 Pascal VOC 데이터셋이 있는 21개 범주인 COCO train 2017의 하위 집합으로 훈련된 사전 훈련 분할 모델이다. S로 표시된 사전 훈련 분할 모델의 보조 특징은 위조된 객체의 위치 정보를 추출하고 디코딩 중 객체 경계 정보의 손실을 방지하는 것을 목표로 한다. 다른 방법보다 최종 예측에서 객체를 더 정확하게 찾는 데 도움이 된다. 분할된 배경과 객체에 대한 정보를 포함하는 보조 특징을 통해 밀집 영역의 분할 결과를 획득할 수 있다. 특징이 연결된 후 특징 어텐션 모듈(Feature Attention Module; FAM)은 채널을 강조하며, 이는 배치 정규화 레이어, ReLU 활성화 함수, 커널 크기가 3×3 인 시그모이드 함수의 컨볼루션 레이어로 구성되며 채널 간 활용 특징 관계를 제공한다. 채널 어텐션은 특징맵의 각 채널이 특징 검출기로 간주되기 때문에 입력 이미지의 의미에 초점을 맞춘다.

수학식 10:

$$\mathbf{P}_i = d_i(\text{FAM}([\text{RB}(\mathbf{x}_{4-i}^{\text{full}}), \mathbf{P}_{i-1}, s_i])),$$

여기서, $\mathbf{P}_{i-1}$ 는 디코더 블록의 단계에 따라 브릿지 또는 이전 디코더로부터의 출력을 나타낸다. 이 경우, s_i 는 위조 이미지 x^{full} 내의 객체들의 위치 정보를 포함하며, 각 디코더 블록에 입력되기 전에, 다른 특징맵 $\mathbf{P}_{i-1}$, $\mathbf{x}_{4-i}^{\text{full}}$ 와의 연결(concatenation)을 위해 특징맵의 해상도를 감소시켜야 한다. 실시예에서는 풀링 연산이 객체의 위치 정보를 왜곡하기 때문에 보조 분할 특징의 해상도를 감소시키기 위해 커널 크기/스트라이프가 4/4, 2/2, 1/1인 컨볼루션 레이어를 적용할 수 있다. 마지막으로, 두 개의 출력, 영역 및 엣지 예측은 두 개의 개별 3×3 컨볼루션 레이어에서 도출될 수 있다. 이미지가 위조되면 수동 크로핑(cropping)이 적용되며, 위조된 이미지의 경계에는 실제 이미지와 구별되는 속성을 가진다. 엣지 스트림으로 이러한 특징을 학습하면 위조 이미지와 진짜 이미지 사이의 특징을 더 깊이 이해하고 이미지의 정교한 표현을 예측하는 방법을 학습하는 데 도움

이 될 수 있다.

수학식 11:

$$\hat{\mathbf{R}} = C_{3 \times 3}(\mathbf{P}_3; w_{\hat{\mathbf{R}}}), \hat{\mathbf{E}} = C_{3 \times 3}(\mathbf{P}_3; w_{\hat{\mathbf{E}}}),$$

여기서, $C_{3 \times 3}$ 은 커널 크기 3×3 , 스트라이드 1, 패딩 1인 컨볼루션 레이어이다. 영역 및 엣지 예측은 독립적인 스트림에서 비롯되기 때문에, 마지막 디코더 블록에는 두 개의 개별 컨볼루션 레이어가 있다. 따라서, $w_{\hat{\mathbf{R}}}$ 와 $w_{\hat{\mathbf{E}}}$ 는 각각 영역 스트림, 엣지 스트림 매개변수이다.

주파수 기반 위조 검출 모델을 위한 다중 작업 학습을 위해 설계된 손실 함수에 대해 설명하기로 한다. 손실 함수는 레이블 마스크 및 마스크 엣지 예측을 위해 구성될 수 있다.

반면에 위조된 영역은 종종 작은 부분으로 나타날 수 있다. 따라서, 모델은 심각한 클래스 불균형으로 인해 주로 음의 영역을 예측하며, 이는 모델 정확도를 크게 저해할 수 있다. 이러한 불균형 문제는 의미 있는 정보 전달을 어렵게 하기 때문에 학습 효율성을 떨어뜨리고, 엣지를 예측할 때 악화된다. 따라서 클래스 불균형 문제를 해결하기 위해 음과 양의 샘플 손실의 가중치를 재조정하기 위해 초점 손실(Focal Loss; FL)을 사용할 수 있다. 훈련 중 집중력이 떨어지면 단순 클래스의 비중이 줄어드는 동시에 도전적 클래스에 대한 관심이 높아진다. 다음과 같이 표현되는 영역 및 엣지 손실을 위해 초점 손실을 사용한다.

수학식 12:

$$\begin{aligned} \mathcal{L}_{FL} = & - \sum_{j=1}^W \sum_{i=1}^H [\alpha(1 - \hat{Y}_{i,j})^\gamma Y_{i,j} \log(\hat{Y}_{i,j}) \\ & + (1 - \alpha)\hat{Y}_{i,j}^\gamma (1 - Y_{i,j}) \log(1 - \hat{Y}_{i,j})], \end{aligned}$$

여기서 $\mathbf{Y} = [Y_{i,j}]$ 와 $\hat{\mathbf{Y}} = [\hat{Y}_{i,j}]$ 는 각각 실측(ground truth), 마스크 엣지에 대한 예측이다. α 는 양의 영역과 음의 영역의 균형을 맞추는 데 사용되는 매개변수이다. γ 는 쉽고 어려운 분류 샘플의 균형을 맞추는 데 사용된다. 예를 들면, 실험을 통해 $\alpha = 0.25$, $\gamma = 2$ 로 설정될 수 있다. 엣지 레이블은 소벨(sobel) 기울기 연산에 의해 획득될 수 있다. $Z(Y)$ 가 레이블 엣지를 나타낼 때 $Z(Y) = \text{Sobel}(Y)$ 이다.

영역 손실을 위해 이진 교차 엔트로피(Binary Cross-Entropy; BCE) 손실을 사용하는데, 다음과 같이 표현될 수 있다.

수학식 13:

$$\begin{aligned} \mathcal{L}_{BCE} = & - \sum_{j=1}^W \sum_{i=1}^H [\hat{Y}_{i,j} \log(Y_{i,j}) \\ & + (1 - \hat{Y}_{i,j}) \log(1 - Y_{i,j})] \end{aligned}$$

Dice 손실은 예측 분할 결과와 정답 주석 사이의 Dice 중첩 계수를 기반으로 네트워크를 최적화하여 전경-배경 불균형을 크게 줄인다. 다음과 같이 표현되는 영역 및 엣지 손실에 대해 Dice 손실을 사용한다.

수학식 14:

$$\mathcal{L}_{Dice} = 2 \frac{\sum_{j=1}^W \sum_{i=1}^H Y_{i,j} \hat{Y}_{i,j}}{\sum_{j=1}^W \sum_{i=1}^H Y_{i,j}^2 + \sum_{j=1}^W \sum_{i=1}^H \hat{Y}_{i,j}^2}$$

마지막으로, 위조 위치 지정(지역화)를 위한 다중 작업 주파수 기반 위조 검출 모델의 총 손실은 다음과 같이

작성될 수 있다.

수학식 15:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{region} + \lambda_2 \mathcal{L}_{edge}$$

여기서, $\mathcal{L}_{region} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice}$ 이고, $\mathcal{L}_{edge} = \mathcal{L}_{FL} + \mathcal{L}_{Dice}$ 이다. FBI-Net 훈련의 목적은 $\mathcal{L}_{total}$ 을 최소화하는 것이며, 이는 각각 $\mathcal{L}_{region}$ 과 $\mathcal{L}_{edge}$ 를 최소화하는 것과 같다. 예를 들면, 실험을 통해 $\lambda_1 = \lambda_2 = 1$ 로 설정하여 네트워크의 위조 검출 능력이 λ_1 과 λ_2 에 의해 크게 영향을 받지 않도록 할 수 있다.

도 4는 일 실시예에 있어서, 위조 위치 검출 시스템의 구성을 설명하기 위한 블록도이고, 도 5는 일 실시예에 있어서, 위조 위치 검출 방법을 설명하기 위한 흐름도이다.

위조 위치 검출 시스템(400)의 프로세서는 이미지 입력부(410) 및 위조 영역 검출부(420)를 포함할 수 있다. 이러한 프로세서의 구성요소들은 위조 위치 검출 시스템에 저장된 프로그램 코드가 제공하는 제어 명령에 따라 프로세서에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다. 프로세서 및 프로세서의 구성요소들은 도 5의 위조 위치 검출 방법이 포함하는 단계들(510 내지 520)을 수행하도록 위조 위치 검출 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서의 구성요소들은 메모리가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다.

프로세서는 위조 위치 검출 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 위조 위치 검출 시스템에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 위조 위치 검출 시스템을 제어할 수 있다. 이때, 프로세서는 이미지 입력부(410) 및 위조 영역 검출부(420) 각각은 메모리에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(510 내지 520)을 실행하기 위한 프로세서의 서로 다른 기능적 표현들일 수 있다.

단계(510)에서 이미지 입력부(410)는 주파수 기반 위조 검출 모델에 입력 이미지를 입력받을 수 있다. 주파수 기반 위조 검출 모델은, 셀프 어텐션을 이용한 다중작업 학습을 통해 주파수를 기반으로 이미지 위조 위치를 검출하도록 구성된 것일 수 있다.

단계(520)에서 위조 영역 검출부(420)는 주파수 기반 위조 검출 모델을 통해 상기 입력받은 입력 이미지와 입력 받은 입력 이미지로부터 주파수 분해를 통해 획득된 저주파 이미지 및 고주파 이미지를 이용하여 입력받은 입력 이미지의 위조 영역을 검출할 수 있다.

이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로

연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

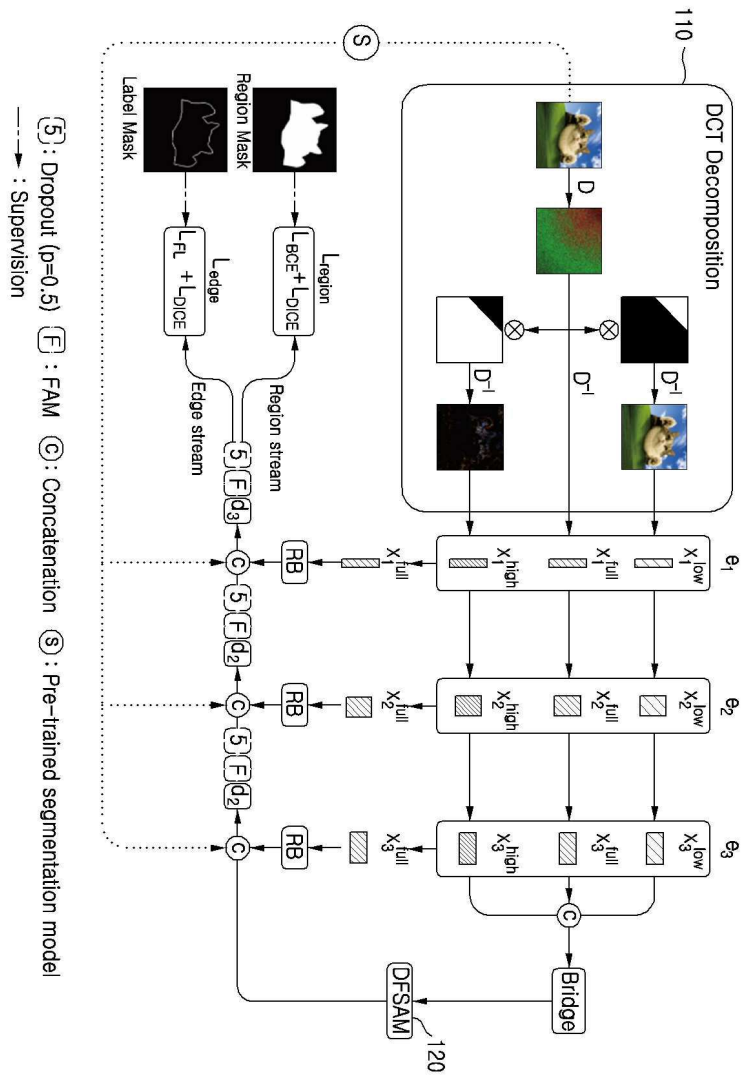
[0099] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0100] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

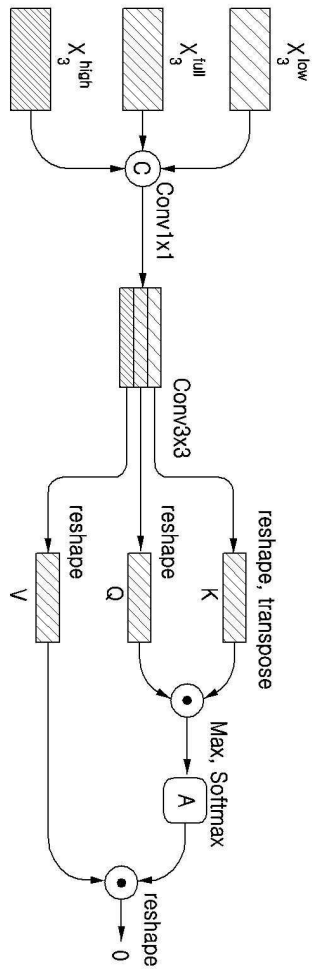
[0101] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

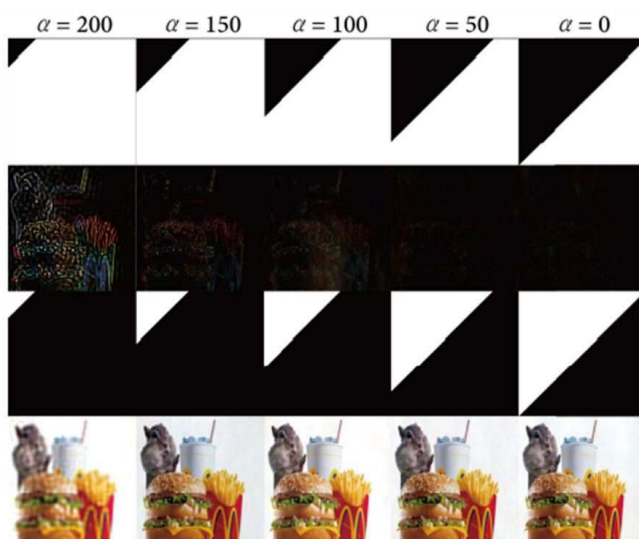
도면1



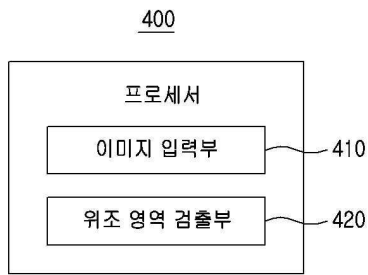
도면2



도면3



도면4



도면5

