



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년12월16일
(11) 등록번호 10-2192211
(24) 등록일자 2020년12월10일

- (51) 국제특허분류(Int. Cl.)
G06T 3/40 (2006.01) G06N 3/08 (2006.01)
- (52) CPC특허분류
G06T 3/4046 (2013.01)
G06N 3/08 (2013.01)
- (21) 출원번호 10-2020-0076546
- (22) 출원일자 2020년06월23일
심사청구일자 2020년06월23일
- (56) 선행기술조사문헌
Chang et al. SYM-parameterized dynamic inference for mixed-domain image translation. IEEE ICCV, 2019년 11월, pp. 4803-4811. 1부.*
Hu et al. Squeeze-and-excitation networks. IEEE CVPR, 2018년 6월, pp. 7132-7141. 1부.*
KR101601660 B1
Kim et al. CANet: A Channel Attention Network to Determine Informative Multi-channel for Image Classification from Brain Signals. IEEE EMBC, 2019년 7월, pp. 680-683. 1부.*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
- (72) 발명자
조근식
인천광역시 연수구 아카데미로312번길 177, 104동 2903호(송도동, 힐스테이트 레이크 송도 1차)
- 김진용
인천광역시 미추홀구 숙골로112번길 11, 501동 803호(도화동, e편한세상도화 5단지)
- (74) 대리인
양성보

전체 청구항 수 : 총 6 항

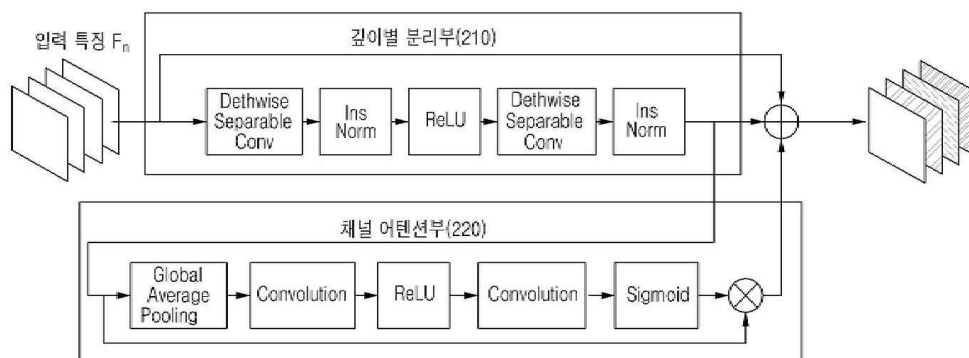
심사관 : 김창원

(54) 발명의 명칭 이미지 변환을 위한 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망

(57) 요약

깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 방법 및 장치가 제시된다. 본 발명에서 제안하는 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 방법은 파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 단계 및 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06T 3/4053 (2013.01)

G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103050
과제번호	2017-0-01642-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	국책과제 운영지원비(산학회계)
연구과제명	[간접비]인공지능을 활용한 콘텐츠 창작 기술-센터운영비(4차년도)
기 여 율	1/1
과제수행기관명	인하대학교 산학협력단
연구기간	2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 단계; 및

깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 단계

를 포함하고,

파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 단계는,

깊이별 컨볼루션의 커널 크기, 폭, 높이, 입력 채널과 특징 맵을 이용하여 깊이별 컨볼루션을 적용하고, 공간적 특징을 다루지 않고 채널에서만 수행되는 포인트 컨볼루션을 적용하고,

깊이별 컨볼루션에서 입력 특징과 출력 특징은 잔차 학습을 진행하는 각 블록에 공급되며, 입력 특징 F_n 의 경우 깊이별 컨볼루션은 하기식과 같이 나타내고,

$$DC(F_n) = F_n \oplus S_{depth}(F_n) \oplus CA(S_{depth}(F_n))$$

깊이별 컨볼루션을 의미하는 $DC(\cdot)$ 는 깊이별 분리 가능 세션(Depthwise Separable Session) $S_{depth}(\cdot)$, 채널 어텐션 세션(Channel Attention Session) $CA(\cdot)$, 및 입력 특징인 F_n 으로 구성되며, 상기 파라미터 모두 잔차 학습(Residual Learning)으로 유기적으로 연결 되어있고, 입력된 F_n 과 $S_{depth}(\cdot)$ 를 통해 출력된 특징인 $S_{depth}(F_n)$ 과 함께 잔차 학습을 진행하며, 추가적으로 $S_{depth}(\cdot)$ 를 통해 추출된 특징을 채널 어텐션 세션에 공급하는 $CA(S_{depth}(F_n))$ 에 대해서도 잔차 학습을 진행하는 이미지 변환 방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 단계는,

채널 간의 상호 의존성을 활용하여 필요한 입력 특징점에 집중하고, 글로벌 에버리지 풀링(Global Average Pooling; GAP)을 사용하여 채널 정보를 압축하고 컨볼루션 층을 통해 입력 특징점을 복원하는

이미지 변환 방법.

청구항 4

제3항에 있어서,

네트워크 내에서 집중해야 할 부분에 대한 입력 특징점들을 강화하고 복원하기 위해 게이팅(gating) 메커니즘을 통해 채널의 통계치를 추출하는

이미지 변환 방법.

청구항 5

파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 깊이별 분리부; 및

깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 채널 어텐션부

를 포함하고,

깊이별 분리부는,

깊이별 컨볼루션의 커널 크기, 폭, 높이, 입력 채널과 특징 맵을 이용하여 깊이별 컨볼루션을 적용하고, 공간적 특징을 다루지 않고 채널에서만 수행되는 포인트 컨볼루션을 적용하고,

깊이별 컨볼루션에서 입력 특징과 출력 특징은 잔차 학습을 진행하는 각 블록에 공급되며, 입력 특징 F_n 의 경우 깊이별 컨볼루션은 하기식과 같이 나타내고,

$$DC(F_n) = F_n \oplus S_{depth}(F_n) \oplus CA(S_{depth}(F_n))$$

깊이별 컨볼루션을 의미하는 $DC(\cdot)$ 는 깊이별 분리 가능 세션(Depthwise Separable Session) $S_{depth}(\cdot)$, 채널 어텐션 세션(Channel Attention Session) $CA(\cdot)$, 및 입력 특징인 F_n 으로 구성되며, 상기 파라미터 모두 잔차 학습(Residual Learning)으로 유기적으로 연결 되어있고, 입력된 F_n 과 $S_{depth}(\cdot)$ 를 통해 출력된 특징인 $S_{depth}(F_n)$ 과 함께 잔차 학습을 진행하며, 추가적으로 $S_{depth}(\cdot)$ 를 통해 추출된 특징을 채널 어텐션 세션에 공급하는 $CA(S_{depth}(F_n))$ 에 대해서도 잔차 학습을 진행하는 이미지 변환 장치.

청구항 6

삭제

청구항 7

제5항에 있어서,

채널 어텐션부는,

채널 간의 상호 의존성을 활용하여 필요한 입력 특징점에 집중하고, 글로벌 에버리지 풀링(Global Average Pooling; GAP)을 사용하여 채널 정보를 압축하고 컨볼루션 층을 통해 입력 특징점을 복원하는

이미지 변환 장치.

청구항 8

제7항에 있어서,

채널 어텐션부는,

네트워크 내에서 집중해야 할 부분에 대한 입력 특징점들을 강화하고 복원하기 위해 게이팅(gating) 메커니즘을 통해 채널의 통계치를 추출하는

이미지 변환 장치.

발명의 설명

기술 분야

[0001] 본 발명은 깊이별 분리 가능한 컨볼루션과 채널 어텐션을 이용하여 효율적인 적대적 생성 신경망을 통해 이미지를 변환하기 위한 방법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 이미지 변환에 대한 연구들이 활발히 연구되고 있다. 이미지 변환은 기하학적 방법, 필터링 등을 이용한 고전적인 방법부터 딥 러닝을 이용한 방법까지 다양한 형태들이 있다. 특히, 딥 러닝을 이용한 방법의 경우 인공신경망의 매우 빠른 발전으로 인해 의미 있는 성과를 거두고 있다. 그 중 이미지 변환 연구에서 가장 큰 성과를 거두고 있는 분야는 적대적 생성 신경망(Generative Adversarial Networks, GANs)를 이용한 연구들이다. 적대적 생성 신경망을 통한 이미지 변환은 이미지의 스타일을 다른 도메인의 이미지형태로 변환하거나(Image-to-Image Translation), 기존 이미지를 고해상도의 새로운 이미지로 생성하거나(Super-resolution), 이미지의 손상된 부분을 채워 넣거나(Image inpainting) 복원하는(Image restoration) 등이 있다. 이는 자율주행 데이터셋 개선, 의료영상 복원 및 개선, 클래스 불균형 데이터셋 증강 등 산업 전반적인 분야에서 사용되고 있다. 그러나, 적대적 생성 신경망은 여러 개의 기존 컨볼루션 신경망으로 이루어져 있고, 이로 인해 다른 인공신경망에 비해 무거운 구조와 많은 계산량을 가지고 있다. 그러므로 딥 러닝 모델의 복잡도(Complexity)를 표현하는 파라미터(parameter)의 수가 굉장히 높게 나타난다. 높은 파라미터의 수는 학습시간과 출력물을 생성하는 추론 시간(Inference time)에 대해 상당한 영향을 줄뿐 아니라 모델 학습, 추론 시 필요로 하는 메모리 리소스가 증가하게 된다. 이러한 문제점들은 현실세계에서 다양한 분야에 이미지 간 변환을 적용하는데 큰 제한점이 된다.

[0004] 도 1은 종래 기술에 따른 기본적인 적대적 생성 신경망의 구조이다.

[0005] 최근 적대적 생성 신경망에 대한 다양하고 많은 양의 연구들이 진행되고 있다. 가장 초기의 모델은 도 1과 같은 형태의 단순한 형태로 진행되었다.

[0006] 무작위 잡음(Random noise)를 입력으로 받은 생성자(Generator)는 무작위 잡음을 통해 가짜 이미지, 다시 말해 생성자를 통해 생성된 이미지(Generated Image)를 생성하기 위한 학습을 한다. 반대로 판별자(Discriminator)는 생성자가 생성한 가짜 이미지와 제공되는 데이터셋의 진짜 이미지(Real Image)를 구분한다. 이 과정에서 생성자는 가짜 이미지가 판별자를 속이기 위해 진짜 이미지와 구별이 가지 않는 이미지를 생성하는 쪽으로 학습되고, 판별자는 가짜 이미지와 진짜 이미지를 구분하기 위한 방향으로 학습된다. 생성자와 판별자는 서로 적대적(Adversarial)인 관계이기 때문에 서로 배반적인 관계이며, 다음과 같은 수식으로 표현된다:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p_z(z)} \log [(1 - D(G(z)))]$$

[0007] 위의 설명과 같이 기본적인 적대적 생성 신경망은 두 개의 네트워크로 되어있다. 기본적인 생성자 한 개와 판별자 한 개로 구성되어 있는 경우도 있지만, 목적에 따라 여러 개의 생성자와 판별자로 구성되어 있다.

[0009] 이미지 간 변환(Image-to-Image Translation)은 서로 다른 도메인의 데이터셋의 스타일을 바꾸는 연구분야다. 다양한 분야에서 쓰일 수 있으며, 최근에는 굉장한 결과물을 내는 연구들이 많이 보여지고 있다. 그러나, 기존 연구에서 사용되는 도메인의 데이터셋은 페어링된(paired) 데이터셋이 주를 이루었다. 이는 데이터셋을 빌드할 때 비용이 많이 든다는 문제점과 현실세계에 존재하는 데이터셋은 대부분 페어링되지 않아 현실에 적용하기 힘들다는 문제점이 있다. 이를 해결하기 위해 페어링되지 않은(unpaired) 데이터셋의 활용을 중점적으로 연구되기 시작했고, 많은 성과들을 이루어 냈다. 종래기술 CycleGAN 에서는 두 도메인간 이미지 데이터셋을 매핑 하고, 순환 일관성 손실(Cycle consistency loss)를 학습 하며 각 도메인을 서로 다른 도메인의 스타일로 이미지 간 변환을 하는 적대적 생성 신경망 모델이 개시되었다. CycleGAN에서, 앞서 말한 페어링되지 않은 데이터셋을 사용하기 때문에 데이터셋을 빌드하는데 있어서 큰 강점을 가지고 있으므로 다양한 분야에서 적용하여 문제를 해결하는 연구들이 보여지고 있다. 일반적인 적대적 생성 신경망과 다르게 두 개의 생성자(G, F)와 2개의 판별자(Dx, Dy)로 이루어져 있으며, 이는 상대적으로 높은 계산량을 보유 하고 있고, 상대적으로 많은 메모리 리소스가 필요하다. 이러한 CycleGAN의 범용성을 이용하여 보다 좋은 품질의 이미지를 생성하기 위한 연구들이 진행되어 왔다.

[0010] 또 다른 종래 기술 AttentionGAN은 CycleGAN을 기반으로 한 적대적 생성 신경망으로 CycleGAN과 마찬가지로 순환 일관성 손실을 이용하여 두 도메인 간의 맵핑을 통하여 이미지 간 변환을 수행하는 모델이다. 추가적으로, 어텐션 메커니즘을 이용했다는 점이 보다 좋은 품질의 이미지를 생성하는데 도움을 주었다. AttentionGAN은 특징점을 같은 네트워크에서 추출하여 다른 두 네트워크에 공급한 다. 콘텐츠 마스크 생성자(Content Mask

Generator)는 입력 이미지에 대한 주요 오브젝트에 대한 특징에 집중한 콘텐츠스 마스크(Contents mask)를 생성한다. 어텐션 마스크 생성자(Attention Mask Generator)에서는 데이터셋에서 중요하게 집중해야 할 부분에 대한 특징을 학습한 어텐션 마스크(Attention mask)를 생성한다. 이렇게 이미지의 특징을 분리(disentangled)하여 각 네트워크는 세부적인 학습을 진행한다. 각 생성자에서 나온 콘텐츠스 마스크와 어텐션 마스크는 정규화 되고 행렬에 대한 곱연산을 진행하여 어텐션 마스크에서 학습한 집중해야 하는 부분에 대해 보전하고 아닌 부분은 지워가는 학습을 진행하게 된다. 이는 모두 AttentionGAN의 내장된 기능으로 진행되므로 보다 효율적인 학습을 할 수 있다. 그러나 CycleGAN과 AttentionGAN은 모두 1:1의 유니모달(unimodal) 출력만 하는 단점이 있다. 결국 한 도메인에서 여러 스타 일의 이미지를 생성하려면 다화자 학습을 해야 하기 때 문에 비효율적이라는 문제점이 있게 된다. 이 문제점을 해결하기 위해, 멀티모달(multimodal)을 출력하는 모델들이 연구되었다. 멀티모달을 중점으로 한 연구들은 앞서 말한 유니모달의 모델보다 상이한 이미지들을 출력하기에 효율적이다.

[0011] 또 다른 종래기술 StarGAN은 한 개의 생성자와 N개의 생성자로 구성되어 있으며, 여기서 N은 만들고자 하는 도메인의 개수를 의미한다. 한 개의 생성자가 입력된 이미지와 도메인 라벨에 대한 학습을 진행하고 해당 라벨의 판별자는 마찬가지로 진짜 데이터(real data)와 가짜 데이터(fake data)를 구분하도록 학습한다. 이렇게 학습하면 기존 유니모달의 모델과는 다르게 한 개의 생성자만으로도 다양한 도메인의 이미지를 학습하여 생성할 수 있기 때문에 효율적이다.

[0012] 또 다른 종래기술 MUNIT은 StarGAN과 다르게 도메인의 라벨없이 스스로 여러 가지의 도메인 이미지를 생성할 수 있다. 이것은 AttentionGAN과 마찬가지로 분리(disentangled)하여 특징점을 추출하고 이것을 콘텐츠스 인코더(Contents encoder)와 스타일 인코더(Style encoder)에 따로 분리하여 주입하기 때문에 가능하다. MUNIT은 마찬가지로 한 개의 생성자와 한 개의 판별자로 이루어져있기 때문에 보다 효율적인 멀티모달 생성이 가능하다.

[0013] 특징의 채널 수가 늘어날수록 딥 러닝 네트워크에서 발생하는 연산이 늘어나고 비용이 증가하게 된다. 이러한 문제를 해결하기 위한 연구들은 계속되어 왔고, 특히 깊이별 분리 컨볼루션을 기반으로 한 연구들은 좋은 성과를 보여주고 있다. 종래기술 Inception과 깊이별 분리 컨볼루션에서 영감을 얻은 Xception에서는 1x1 전환을 먼저 수행한 다음 모든 출력 채널에 대해 공간 상관 맵핑을 개별적으로 수행하는 극한 버전 시작 모듈을 제안했다.

[0014] MobileNet은 깊이 분리 가능한 컨볼루션과 입력 및 출력 채널을 제어하기 위한 폭승수와 입력 이미지의 크기를 조정하기 위한 해상도 승수로 구성된 추가 수축 하이퍼 파라미터를 사용한다. ShuffleNet은 포인트 단위 컨볼루션(Pointwise Convolution)이 여전히 높은 비용 영역임을 강조한다. 이 문제를 해결하기 위해 ShuffleNet 은 모든 가중치를 연결하지 않은 채널 스페어스(Channel sparse)를 설계하고 그룹을 섞어 특정 영역에 대한 정보 흐름만 입력으로 얻는 문제를 방지했다.

[0015] 기존 어텐션 메커니즘은 자연어처리(Natural Language Process) 분야에서 주로 사용되었지만, 최근에는 컴퓨터 비전 분야에서도 다양하게 사용되고 있다. Image Classification, Super-Resolution, Image Detection 등 다양하게 사용되고 있으며, 다양한 종 류의 어텐션 메커니즘이 연구되고 있다. SeNet에서 사용되는 채널 어텐션(Channel Attention)은 채널 정보를 압축하고 리스케일링 하여 채널 간의 의존도를 분석 하고 집중할 부분에 더 높은 가중치를 주어 성능을 개선한다. 또한, CBAM은 채널뿐만 아니라 공간적 어텐션(Spatial Attention) 메커니즘을 결합하여 공간정보를 추가적으로 스케일링하여 성능을 개선하였다.

발명의 내용

해결하려는 과제

[0016] 본 발명이 이루고자 하는 기술적 과제는 페어링되지 않은 이미지 간 변환(Unpaired Image-to-Image Translation)에서 일반적인 컨볼루션을 깊이별 분리가능한 컨볼루션(Depthwise separable convolution)으로 대체하고 채널 어텐션을 적용하는 DCBlock 모듈을 제공하고, DCBlock을 통해 메모리 리소스가 제한된 환경에서 사용할 수 있도록 학습 파라미터의 수를 줄이고자 한다.

과제의 해결 수단

[0017] 일 측면에 있어서, 본 발명에서 제안하는 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 방법은 파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 단계 및 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용

간의 균형을 맞추기 위해 채널 어텐션을 적용하는 단계를 포함한다.

[0018] 파라미터의 수를 줄이기 위해 깊이 분리가능한 컨볼루션을 적용하는 단계는 깊이별 컨볼루션의 커널 크기, 폭, 높이, 입력 채널과 특징 맵을 이용하여 깊이별 컨볼루션을 적용하고, 공간적 특징을 다루지 않고 채널에서만 수행되는 포인트 컨볼루션을 적용한다.

[0019] 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 단계는 채널 간의 상호 의존성을 활용하여 필요한 입력 특징점에 집중하고, 글로벌 에버리지 풀링(Global Average Pooling; GAP)을 사용하여 채널 정보를 압축하고 컨볼루션 층을 통해 입력 특징점을 복원하며, 네트워크 내에서 집중해야 할 부분에 대한 입력 특징점들을 강화하고 복원하기 위해 게이팅(gating) 메커니즘을 통해 채널의 통계치를 추출한다.

[0020] 또 다른 본 발명에서 제안하는 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 장치는 파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 깊이별 분리부 및 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 채널 어텐션부를 포함한다.

발명의 효과

[0021] 본 발명의 실시예들에 따른 DCBlock은 표준 컨볼루션(Standard Convolution)보다 상대적으로 적은 계산 비용을 갖는 깊이별 분리가능한 합성곱(Depthwise Separable Convolution)과 깊이별 분리 가능한 합성곱으로 인해 발생하는 정보손실(Information loss)을 보완하기 위한 채널 어텐션(Channel Attention)을 적용하여 계산량을 표현하는 지표인 파라미터(parameter)의 수를 최대 91.6% 줄임에도 불구하고 종래기술의 모델들과 비교하여 유사한 품질의 이미지를 생성하는 성능을 보여준다.

도면의 간단한 설명

[0022] 도 1은 종래 기술에 따른 기본적인 적대적 생성 신경망의 구조이다.

도 2는 본 발명의 일 실시예에 따른 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 장치의 구조를 나타내는 도면이다.

도 3은 본 발명의 일 실시예에 따른 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 방법을 설명하기 위한 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 깊이별 분리 가능한 컨볼루션 구성을 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 채널 어텐션 메커니즘을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0023] 본 발명의 실시예에 따르면, 학습 파라미터의 수를 줄이기 위해 일반적인 컨볼루션을 깊이별 분리 가능한 컨볼루션(depthwise separable convolution)으로 대체했다. 그러나, 깊이별 분리가능한 컨볼루션은 파라미터의 수는 줄어들지만 그만큼 정보 손실(Information loss)을 발생시킨다. 정보 손실은 적대적 생성 신경망 모델이 좋지 않은 퀄리티의 이미지 생성을 하는 것에 결정적인 영향을 끼친다.

[0024] 학습 파라미터의 수를 줄이기 위해 깊이별 분리 가능한 컨볼루션을 적용하는 경우 발생하는 정보 손실을 보상하기 위해, 본 발명에서는 잔차 블록(Residual block)에 채널 어텐션을 적용한다. 채널 어텐션을 적용하면 중요한 특징점에 집중을 하게 되고 정보 손실에 대한 완화작용을 하게 되어 이미지의 품질에 대한 보장을 할 수 있다.

[0025] 따라서, 본 발명에서는 앞서 말한 기술들을 적용하여 DCBlock을 제안한다. DCBlock은 일반적인 적대적 생성 신경망에서 사용되는 "잔차 블록(Residual Block)"을 대체하여 이미지간 변환을 실행하는 구조로 되어있다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0027] 도 2는 본 발명의 일 실시예에 따른 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 장치의 구조를 나타내는 도면이다.

[0028] 본 발명에서는 적대적 생성 신경망을 이용해 페어링되지 않은 이미지 간 변환(Unpaired Image-to-Image Translation) 시 학습 모델이 무거워지는 문제를 해결하기 위해 DCBlock (Depthwise separable Channel

attention Block)을 제안한다.

[0029] 제안하는 모듈 DCBlock은 도 2에서 볼 수 있듯, 깊이별 분리부(210)(다시 말해, 깊이별 분리 가능 세션(Depthwise Separable Session))과 채널 어텐션부(220)(다시 말해, 채널 어텐션 세션(Channel Attention Session))의 두 세션으로 구성된다.

[0030] 깊이별 분리부(210)는 파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용한다.

[0031] 깊이별 분리부(210)는 깊이별 분리 컨볼루션(Dethwise Separable Conv) 블록, Ins Norm(Instance Normalization) 블록, ReLU 블록, 깊이별 분리 컨볼루션(Dethwise Separable Conv) 블록, Ins Norm(Instance Normalization) 블록을 거쳐, 깊이 컨볼루션의 커널 크기, 폭, 높이, 입력 채널과 특징 맵을 이용하여 깊이별 컨볼루션을 적용하고, 공간적 특징을 다루지 않고 채널에서만 수행되는 포인트 컨볼루션을 적용한다.

[0032] 깊이 분리가능한 컨볼루션을 적용하여 파라미터의 수를 줄이는 경우 정보 손실이 발생한다. 이러한 정보손실을 잔류 연결로 극복할 수 있지만 이미지 간 변환에서 이를 완화하지 않아 모델의 결과물이 기존 모델의 이미지 품질만큼 나타나지 않는다. 정보손실은 각 채널에 대해 깊이 분리 가능한 컨볼루션이 수행되므로 전체 부분에 나타나는 특징의 손실은 불가피하기 때문에 일어난다. 그러므로 출력된 이미지의 품질과 계산 비용 간의 최적 균형을 맞추기 위해 정보 손실을 보상하는 기술을 추 가해야 했다. 따라서 필요한 정보를 최대한 유지하고 심 층 구조에서도 정보를 잃지 않도록 채널 어텐션을 적용 한다.

[0033] 채널 어텐션부(220)는 GAP(Global Average Pooling) 블록, 컨볼루션(Convolution) 블록, ReLU 블록, 컨볼루션(Convolution) 블록, 시그모이드(Sigmoid) 블록을 거쳐, 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용한다.

[0034] 채널 어텐션부(220)는 채널 간의 상호 의존성을 활용하여 필요한 입력 특징점에 집중하고, 글로벌 에버리지 풀링(Global Average Pooling; GAP)을 사용하여 채널 정보를 압축하고 컨볼루션 층을 통해 입력 특징점을 복원한다. 그리고, 네트워크 내에서 집중해야 할 부분에 대한 입력 특징점들을 강화하고 복원하기 위해 게이팅(gating) 메커니즘을 통해 채널의 통계치를 추출한다.

[0035] DCBlock에서 입력 특징과 출력 특징은 잔차 학습을 진행하는 각 블록에 공급된다. 입력 특징 F_n 의 경우 DCBlock은 다음과 같이 공식화할 수 있다:

$$[0036] \quad DC(F_n) = F_n \oplus S_{depth}(F_n) \oplus CA(S_{depth}(F_n))$$

[0037] DCBlock 모듈을 의미하는 $DC(\cdot)$ 는 깊이별 분리 가능 세션(Depthwise Separable Session) $S_{depth}(\cdot)$, 채널 어텐션 세션(Channel Attention Session) $CA(\cdot)$, 그리고 입력 특징인 F_n 으로 구성되어 있다. 이는 모 두 잔차 학습(Residual Learning)으로 유기적인 연결이 되어있다. 입력된 F_n 과 이를 $S_{depth}(\cdot)$ 를 통해 출력된 특징인 $S_{depth}(F_n)$ 과 함께 잔차 학습을 진행한다. 추가적으로 $S_{depth}(\cdot)$ 를 통해 추출된 특징을 채널 어텐션 세션에 공급해준 $CA(S_{depth}(F_n))$ 도 잔차 학습을 진행한다. 이에 해당되는 부분은 다음과 같이 공식화 할 수 있다:

$$[0038] \quad CA(S_{depth}(F_n)) = S_{depth}(F_n) \otimes S_{*}(S_{depth}(F_n))$$

[0039] 채널 어텐션으로부터 추출된 특징점인 $S_{*}(S_{depth}(F_n))$ 은 $S_{depth}(F_n)$ 과 곱연산을 진행하게 되고 게이팅(gating) 메커니즘으로 인해 특징에는 자연스럽게 채널의 통계치로 얻어진 집중점에 대해 입혀지게 되어 중요한 부분의 정보가 강화되고 정보 손실을 방지할 수 있다. 이 방법을 통 해 DCBlock은 이미지 간 변환 시 많은 수의 파라미터를 필요로 하는 학습모델의 파라미터 수를 줄일 수 있고, 기 존 이미지 간 변환을 학습하는 모델과 유사한 품질의 이미지를 생성할 수 있게 된다.

[0041] 도 3은 본 발명의 일 실시예에 따른 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 방법을 설명하기 위한 흐름도이다.

[0042] 제안하는 깊이별 분리가능한 컨볼루션과 채널 어텐션을 이용한 효율적인 적대적 생성 신경망의 이미지 변환 방법은 파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용하는 단계(310) 및 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용하는 단계(320)를 포함한다.

[0043] 단계(310)에서, 파라미터의 수를 줄이기 위해 깊이 분리 가능한 컨볼루션을 적용한다.

[0044] 깊이 분리가능한 컨볼루션을 적용하여 파라미터의 수를 줄이는 경우 정보 손실이 발생한다. 이러한 정보손실을 잔류 연결로 극복할 수 있지만 이미지 간 변환에서 이를 완화하지 않아 모델의 결과물이 기존 모델의 이미지 품질만큼 나타나지 않는다. 정보손실은 각 채널에 대해 깊이 분리 가능한 컨볼루션이 수행되므로 전체 부분에 나타나는 특징의 손실은 불가피하기 때문에 일어난다. 그러므로 출력된 이미지의 품질과 계산 비용 간의 최적 균형을 맞추기 위해 정보 손실을 보상하는 기술을 추가해야 했다. 따라서 필요한 정보를 최대한 유지하고 심층 구조에서도 정보를 잃지 않도록 채널 어텐션을 적용 한다.

[0045] 단계(320)에서, 깊이 분리 가능한 컨볼루션을 적용하여 발생하는 정보 손실을 보상하여 출력된 이미지의 품질과 계산 비용 간의 균형을 맞추기 위해 채널 어텐션을 적용한다. 아래에서 도 4 및 도 5를 참조하여 더욱 상세히 설명한다.

[0047] 도 4는 본 발명의 일 실시예에 따른 깊이별 분리 가능한 컨볼루션 구성을 설명하기 위한 도면이다.

[0048] 앞서 설명했듯, 본 발명에서는 파라미터의 수를 줄이기 위해 도 4(a)와 같은 깊이별 컨볼루션(Depthwise Convolution)과 도 4(b)와 같은 포인트 컨볼루션(Pointwise Convolution)으로 이루어진 깊이별 분리 가능한 컨볼루션을 사용한다.

[0049] 깊이 컨볼루션에는 공간 피처를 추출하기 위한 채널 수에 대한 필터가 있으므로 입력 및 출력 채널 수가 동일하다. 깊이 컨볼루션은 다음과 같이 나타낼 수 있다:

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \times F_{k+i-1,l+j-1,m}$$

[0051] 여기서 $\hat{K}$ 은 깊이 컨볼루션의 커널 크기를 나타내고, i, j, m은 폭, 높이, 입력 채널을 나타내고 F 는 특징 맵을 나타낸다. 포인트 컨볼루션은 1x1 컨볼루션이며 필터의 크기는 1x1로 고정된다. 깊이 컨볼루션과 달리, 포인트 컨볼루션은 공간적 특징을 다루지 않고 채널에서만 수행된다. 이것은 DCBlock에서 계산량을 크게 줄이는데 도움이 된다. 표 1은 두 변환의 파라미터의 수와 계산 비용의 차이를 보여준다.

[0052] <표 1>

Method	Standard Convolution	Depthwise Separable Convolution
# of Param	$K^2 \times C \times M$	$C \times (K^2 + M)$
Computational cost	$K^2 \times C \times M \times H \times W$	$WC \times H \times W \times (K^2 + M)$

[0054] 여기서 K 는 커널 크기, C, M 은 입력 및 출력 채널 크기를 각각 나타내고 H, W 는 입력 높이, 너비를 각각 나타낸다. 표 1에 나타난 바와 같이, 표준 컨볼루션은 계산 비용이 $K^2 \times C \times M \times H \times W$ 이고 깊이 분리가능

한 컨볼루션은 $C \times H \times W \times (K^2 + M)$ 이다. 차이를 보기 위해 두 비용을 나누면 비용이 $\frac{1}{M} + \frac{1}{K^2}$ 감소한 것으로 나타난다. 이는 상대적으로 깊은 모델이 사용되는 적대적 생성 신경망에서 비례 적으로 증가하는 파라미터의 수를 줄이는 데 큰 효과가 있다.

[0056] 도 5는 본 발명의 일 실시예에 따른 채널 어텐션 매커니즘을 설명하기 위한 도면이다.

[0057] 채널 어텐션 메커니즘은 채널 간의 상호 의존성을 활용하여 필요한 특징점에 집중하는 어텐션의 변형 중

하나다.

[0058] 도 5에서 볼 수 있듯, GAP(Global Average Pooling) 블록, FC(Fully connected Layer)(다시 말해, 컨볼루션(Convolution)) 블록, ReLU 블록, FC(Fully connected Layer)(다시 말해, 컨볼루션(Convolution)) 블록, 시그모이드(Sigmoid) 블록을 거쳐, 입력 특징(510)에 채널 어텐션(520)을 적용하여, 글로벌 에버리지 풀링(Global Average Pooling: GAP)을 사용하여 채널 정보를 압축하고 컨볼루션 층을 통해 특징점을 복원한다. 다음 식은 GAP 를 통한 입력 특징점 $z_c(x_c)$ 을 구하는 식이다:

$$z_c(x_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j)$$

[0059]

[0060] 이어서, 채널의 통계치는 게이팅(gating) 메커니즘을 통해 추출된다. 이 효과로 네트워크 내에서 집중해야 할 부분에 대한 특징들이 강화되고 복원된다. 따라서 심층모델에서도 가능한 많은 특징점을 유지하고 깊이 분리가 가능한 컨볼루션으로 인한 정보 손실을 해결하기 위해 채널 어텐션을 사용하여 복원된 입력 특징점을 출력(530)한다.

[0062]

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, FPA(field programmable 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.

[0063]

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(embodiment)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.

[0064]

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.

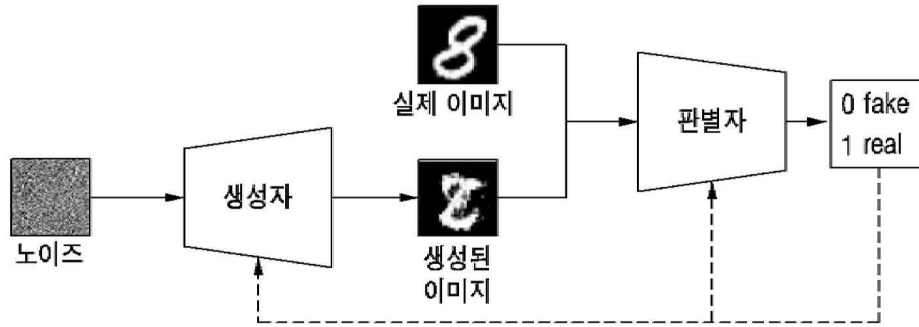
[0065]

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.

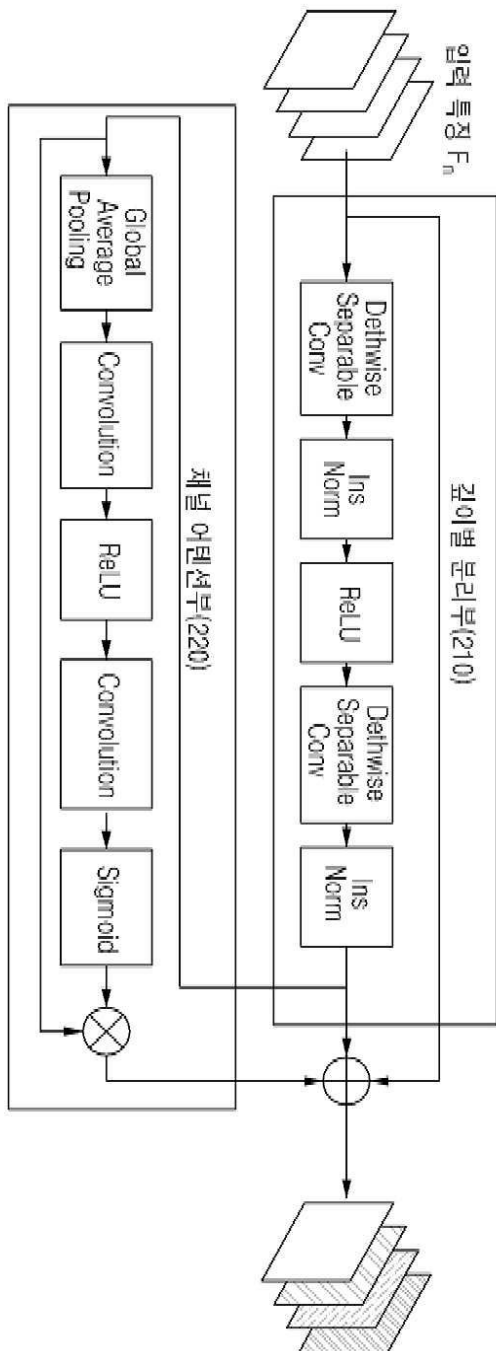
[0066] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

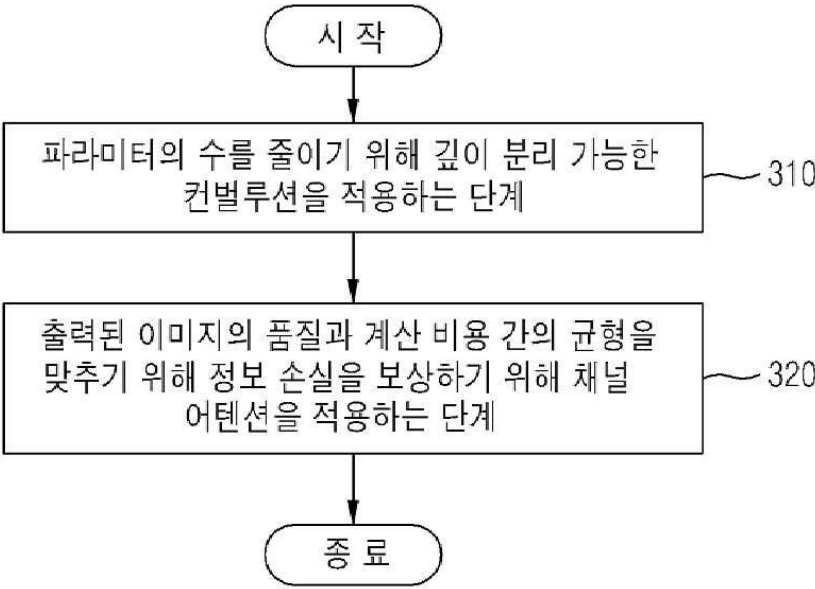
도면1



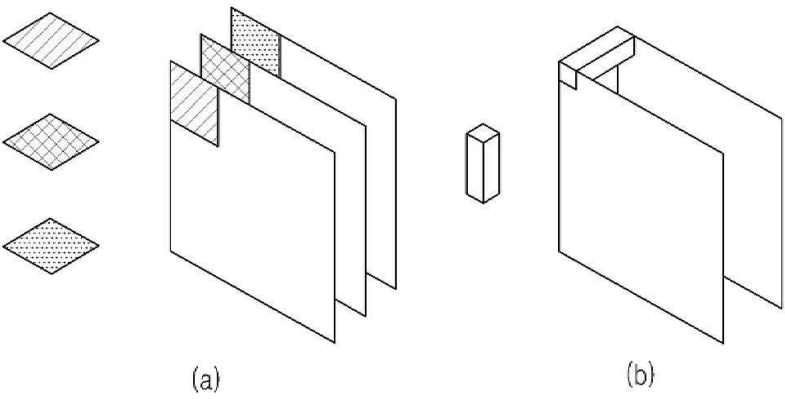
도면2



도면3



도면4



도면5

