Generative adversarial network (GAN) in medical data

한림대학교 인공지능융합학부(과)

원동옥

소 개

Dong-Ok Won, Ph.D.

2012-2019 - Integrated M.S. and Ph.D., Dept. of Brain and Cognitive Engineering, Korea University
Pattern Recognition & Machine Learning Lab., Supervisor: Prof. Seong-Whan Lee

2019-2020 - Research Professor, Dept. of Brain and Cognitive Engineering, Korea University

2020~현재 - Assistant Professor, School of AI Convergence, Hallym University

Research areas

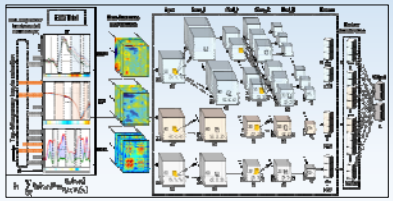
Artificial intelligence, machine learning, human-machine interaction, computational neuroscience, big data analysis

Presentation and demonstration in **NeurIPS 2019** and **IJCAI 2018**

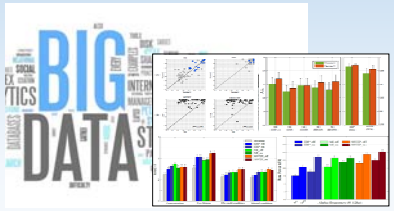
Best Research Award in Microsoft Research Asia Academic Day 2019

Collaboration with Prof. K.-R Müller (Technical University of Berlin/ Google Brain)

주요 연구 분야



Deep learning and reinforcement learning



Machine learning and big data analysis

Human-machine interaction
(Brain-computer interface)

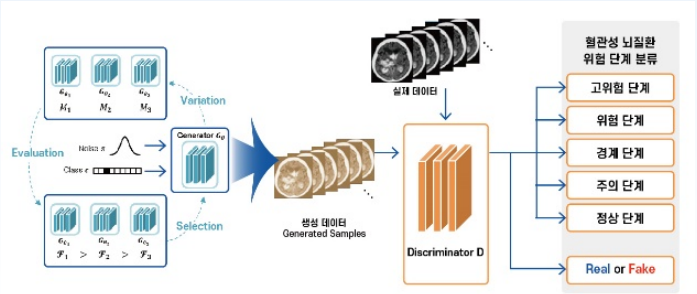
Artificial intelligence

주요 관심 분야 - 뇌혈관질환 전주기 관리 플랫폼

주요 관심 분야 - 뇌혈관질환 전주기 관리 플랫폼

GAN (Generative Adversarial Network) - 의료 인공지능 적용

- 딥러닝 모델 중 이미지 생성에 널리 쓰이는 모델
  - GAN (Generative Adversarial Network)
  - 절대적으로 부족한 의료 데이터 보강을 위한 의료데이터(영상) 생성적 적대 신경망 기술

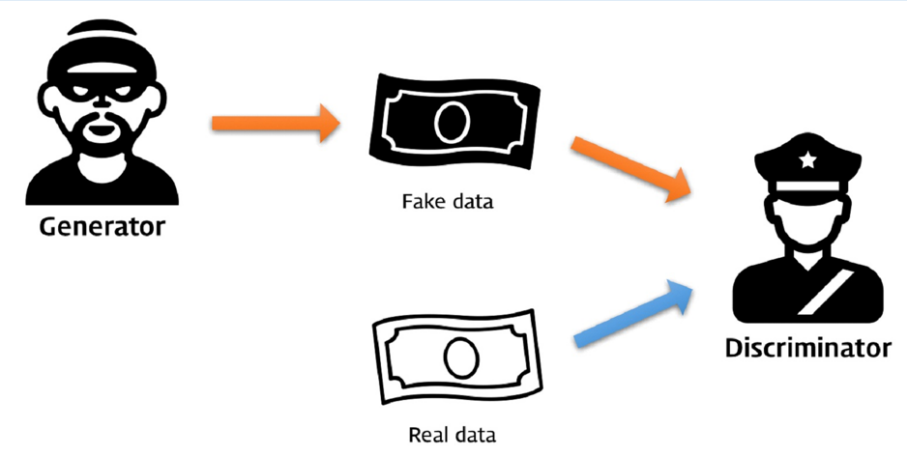


생성적 적대 신경망 기반 의료 가상 데이터 생성 기술 예시

GAN (Generative Adversarial Network) (1/6)

영화 catch me if you can (2002)

GAN (2/6)



https://dreamgonfly.github.io/blog/gan-explained/

GAN (3/6)

- Generative, GAN이 생성(Generation) 모델이라는 것을 뜻
  - 생성 모델이란 '그럴듯한 가짜'를 만들어내는 모델
  - 언뜻 보면 진짜 같은 가짜 사람 얼굴 사진을 만들어내거나 실제로 있을 법한 고양이 사진을 만들어내는 것이 생성 모델의 예

'그럴듯하다'

  - 비교적 쉬운 예. 키와 몸무게의 생성 모델
  - 어려운 예. 사람얼굴 or 고양이 사진 등 (256 x 256 x 3(RGB))

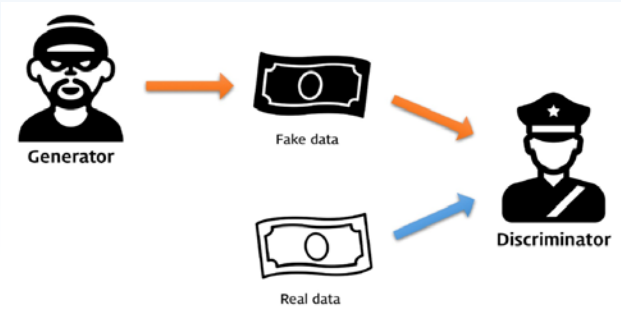
https://dreamgonfly.github.io/blog/gan-explained/

GAN (4/6)

- Adversarial, GAN이 두 개의 모델을 적대적(Adversarial)으로 경쟁시키며 발전시킨다는 것을 뜻
  - 위조지폐범과 경찰을 생각해보자. 이 둘은 적대적인 경쟁 관계!

    위조지폐범은 경찰을 속이기 위해 점점 지폐 위조 제조 기술을 발전시키고,
    경찰은 위조지폐범을 잡기 위해 점점 위폐를 찾는 기술을 발전시킨다.

    시간이 흐르면 위조지폐범의 위폐 제조 기술은 완벽에 가깝게 발전할 것



https://dreamgonfly.github.io/blog/gan-explained/

GAN (5/6)

- Adversarial, GAN이 두 개의 모델을 적대적(Adversarial)으로 경쟁시키며 발전시킨다는 것을 뜻
  - 위조지폐범에 해당하는 생성자(Generator),

    경찰에 해당하는 구분자(Discriminator)를 경쟁적으로 학습
  - 생성자의 목적은 그럴듯한 가짜 데이터를 만들어서 구분자를 속이는 것

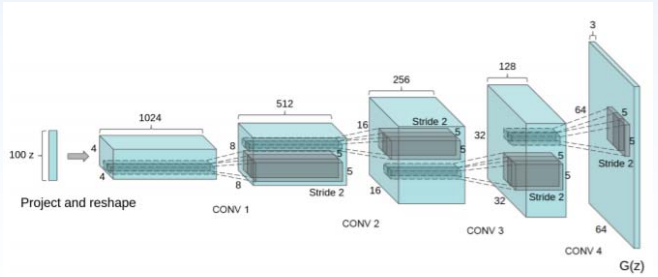
    구분자의 목적은 생성자가 만든 가짜 데이터와 진짜 데이터를 구분하는 것
  - 이 둘을 함께 학습시키면서 진짜와 구분할 수 없는 가짜를 만들어내는 생성자를 얻을 수 있음

이것이 GAN의 핵심적인 아이디어인
적대적 학습(Adversarial Training)

https://dreamgonfly.github.io/blog/gan-explained/

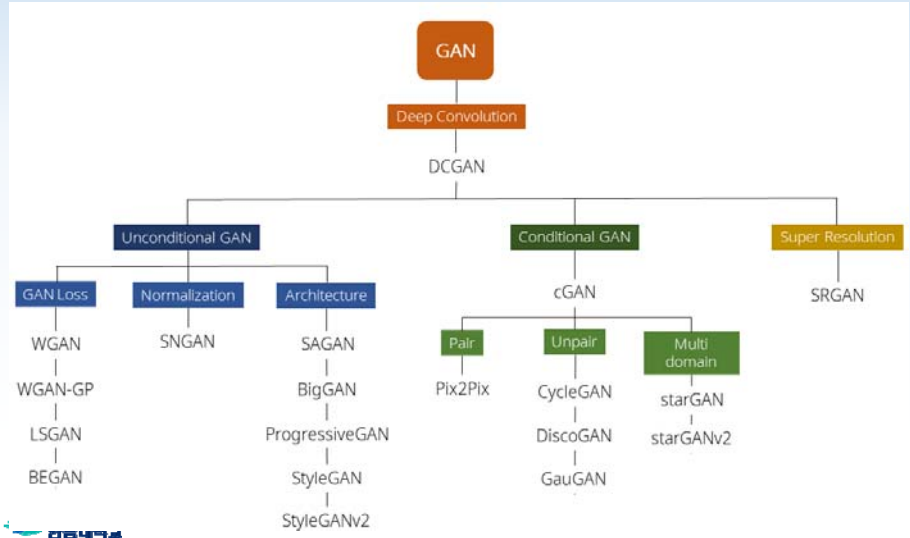
GAN (6/6)

- Network, 인공신경망(Artificial Neural Network) or 딥러닝(Deep Learning)으로 만듦
  - 적대적 학습이라는 개념을 구현하기 위해 반드시 딥러닝을 써야 하는 것은 아니지만,
  - 딥러닝은 강력한 머신러닝 모델을 가능하게 만드는 기술



https://dreamgonfly.github.io/blog/gan-explained/

GAN 종류



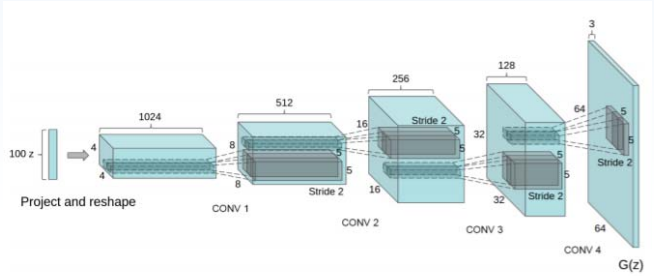
https://ysbsb.github.io/gan/2020/06/17/GAN-newbie-guide.html

GAN 종류 (Cont.)

- **Generative Adversarial Nets [NIPS 2015]**
  - Ian Goodfellow가 최초로 GAN (Generative Adversarial Nets)를 제안한 논문.
  - 새로운 이미지를 생성하는 생성자 Generator와 샘플 데이터와 생성자가 생성한 이미지를 구분하는 구별자 Discriminator 두 개의 네트워크 구조를 제안
  - 생성자는 구별자를 속이면서 이미지를 잘 생성하려고 하며, 구별자는 주어진 이미지가 진짜인지 가짜인지 판별

GAN 종류 (Cont.)

- **GAN의 전성시대를 연 DCGAN [ICLR 2016]**
- **Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks**
  - 기존의 GAN에 CNN의 네트워크를 도입한 DCGAN (Deep Convolutonal Generative Adversarial Networks) 제안
  - GAN은 학습이 불안정하기로 악명이 높다.
  - GAN만 사용하면 성능이 좋지 않았으나, DCGAN 이후로부터 GAN의 발전이 많이 됨
    - ✓ 선형 레이어와 풀링 레이어(Pooling Layer)를 최대한 배제 (이미지 위치정보 잃기 때문)
    - ✓ 합성곱(Convolution)과 'Transposed Convolution(Fractional-Strided Convolution)'으로 네트워크 구조



GAN 종류 (Cont.)

- **cGAN(Conditional GAN)**
  - 이미지를 처음부터 생성하기보다 이미 있는 이미지를 다른 영역의 이미지로 변형하고 싶은 경우
  - 스케치에 채색하거나, 흑백 사진을 컬러로 만들거나, 낮 사진을 밤 사진으로 바꾸고 싶을 때
  - 기존의 GAN의 생성자가 랜덤 벡터를 입력으로 받는 것에 비해 cGAN의 생성자는 변형할 이미지를 입력
  - 그 뒤 생성자는 입력 이미지에 맞는 변형된 이미지를 출력
    - ✓ Ex. 스케치 사진을 받은 생성자는 그 스케치에 맞는 색을 칠한 뒤 채색된 이미지를 출력
    - ✓ 구분자는 스케치와 채색된 이미지를 모두 보고 그 채색된 이미지가 과연 스케치에 어울리는지 판단

구분자를 속이기 위해서
생성자는 '첫째, 진짜 같은 이미지를 만들어야 하고 둘째, 스케치에 맞는 이미지를 만들어야 한다.

다양한 종류의 GAN

GAN: https://arxiv.org/abs/1406.2661
DCGAN: https://arxiv.org/abs/1511.06434
cGAN: https://arxiv.org/abs/1611.07004
WGAN: https://arxiv.org/abs/1701.07875
EBGAN: https://arxiv.org/abs/1609.03126
BEGAN: https://arxiv.org/abs/1703.10717
CycleGAN: https://arxiv.org/abs/1703.10593
DiscoGAN: https://arxiv.org/abs/1703.05192
StarGAN: https://arxiv.org/abs/1711.09020
SRGAN: https://arxiv.org/abs/1609.04802
SEGAN: https://arxiv.org/abs/1703.09452

PlausMal-GAN: Plausible Malware Training Based on Generative Adversarial Networks for Analogous Zero-day Malware Detection

https://www.computer.org/csdl/journal/ec/5555/01/09767588/1D4SjLGhJT2

PlausMal-GAN

IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, VOL. XX, NO. XX, XX 2022

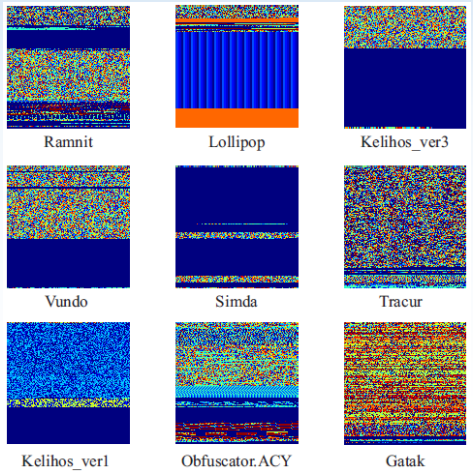
PlausMal-GAN: Plausible Malware Training Based on Generative Adversarial Networks for Analogous Zero-day Malware Detection

Dong-Ok Won, Yong-Nam Jang, and Seong-Whan Lee, *Fellow, IEEE*

Abstract—Zero-day malicious software (malware) refers to a previously unknown or newly discovered software vulnerability. The fundamental objective of this paper is to enhance detection for analogous zero-day malware by efficient learning to plausible generated data. To detect zero-day malware, we proposed a malware training framework based on the generated analogous malware data using generative adversarial networks (PlausMal-GAN). Thus, the PlausMal-GAN can suitably produce analogous zero-day malware images with high quality and high diversity from the existing malware data. The discriminator, as a detector, learns various malware features using both real and generated malware images. In terms of performance, the proposed framework showed higher and more stable performances for the analogous zero-day malware images, which can be assumed to be analogous zero-day malware data. We obtained reliable accuracy performances in the proposed PlausMal-GAN framework with representative GAN models (i.e., deep convolutional GAN, least-squares GAN, Wasserstein GAN with gradient penalty, and evolutionary GAN). These results indicate that the use of the proposed framework is beneficial for the detection and prediction of numerous and analogous zero-day malware data from noted malware when developing and updating malware detection systems.

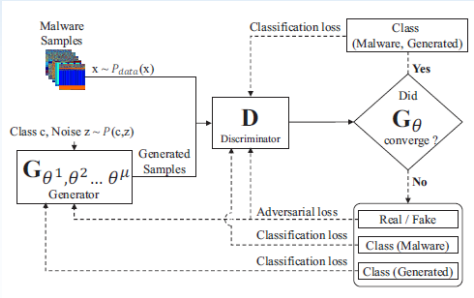
PlausMal-GAN

- **Examples of malware images**



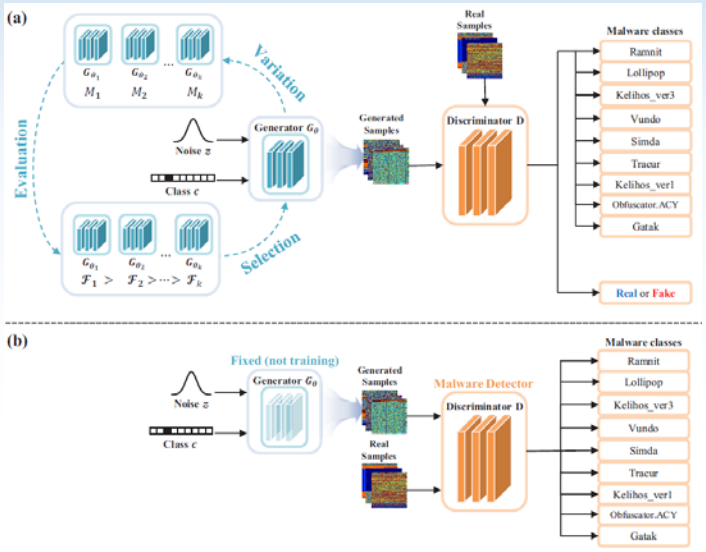
PlausMal-GAN

- The architectures of the proposed framework for analogous zero-day malware detection



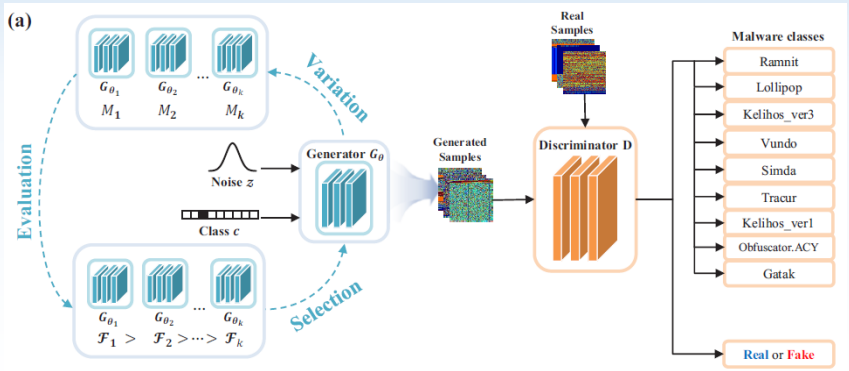
PlausMal-GAN

- The proposed PlausMal-GAN framework consists of two-phases



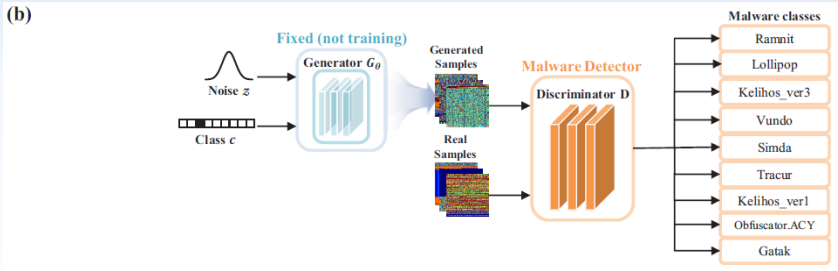
PlausMal-GAN

- The generator and discriminator training based on GAN with malware classifier



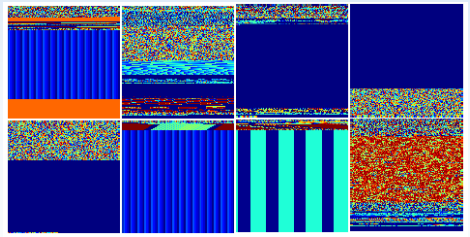
PlausMal-GAN

- Training the discriminator as a zero-day malware detector from plausible malware augmentation

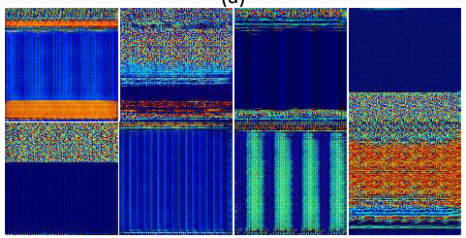


PlausMal-GAN

- **Examples of (a) real malware images and (b) generated malware images in the proposed framework.**



(a)



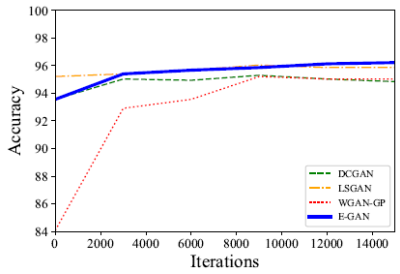
(b)

PlausMal-GAN

- **Results**

Comparison of malware classification accuracies in the proposed framework with four representative GAN models and previous methods

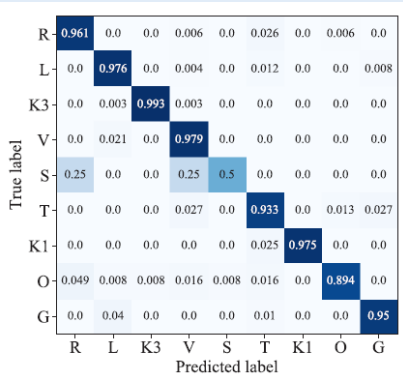
Model	MLP	CNN	GAN	tGAN	Proposed Framework (PlausMal-GAN) DCGAN	LSGAN	WGAN-GP	E-GAN
Accuracy (%)	83.06	94.63	87.81	88.10	94.99	96.02	94.86	**96.35**
Std. dev.	7.54e-04	2.12e-05	3.44e-05	8.05e-05	0.596	0.351	0.255	0.539



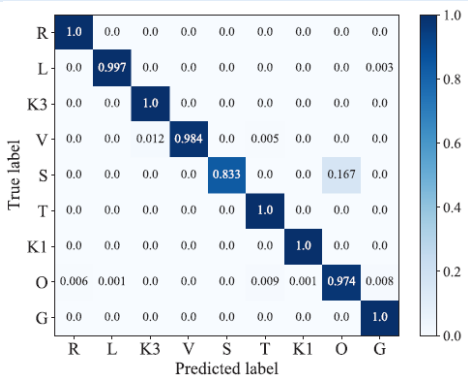
Classification accuracy according to the training iterations for the proposed framework with four representative models.

PlausMal-GAN

- **Results**



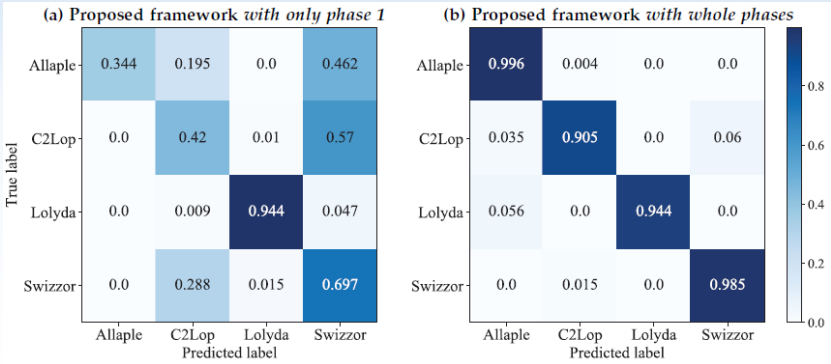
Confusion matrix for malware classification results in 9:1 train-test ratio



Confusion matrix for malware classification results in 5:5 train-test ratio

PlausMal-GAN

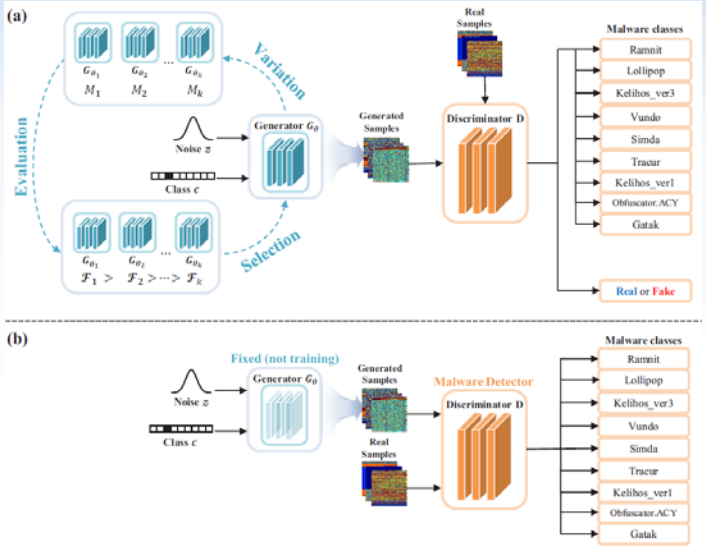
- **Results**



Confusion matrix for zero-day malware classification results (session B) in the proposed framework (E-GAN) with (a) only phase 1 and (b) whole phases (1&2)

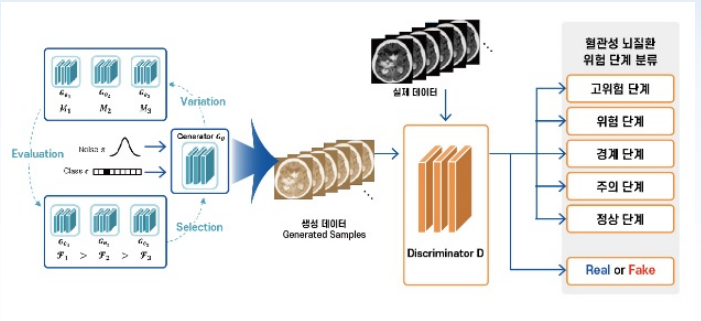
PlausMal-GAN

- The proposed PlausMal-GAN framework



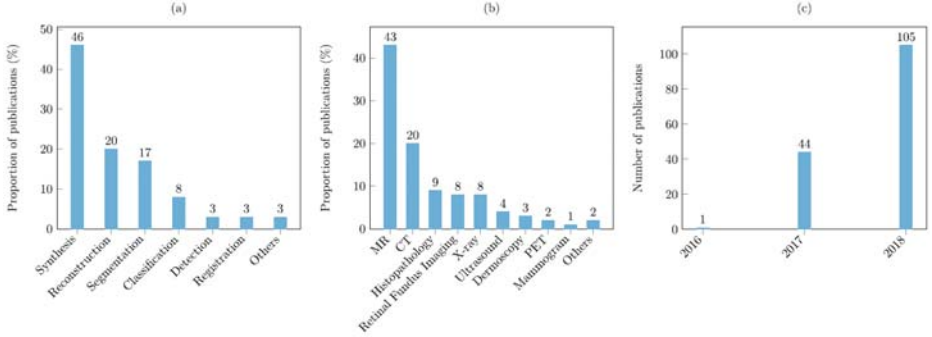
의료 인공지능 적용

- GAN Framework
  - ➢ 절대적으로 부족한 의료 데이터 보강을 위한 의료데이터(영상) 생성적 적대 신경망 기술



생성적 적대 신경망 기반 의료 가상 데이터 생성 기술 예시

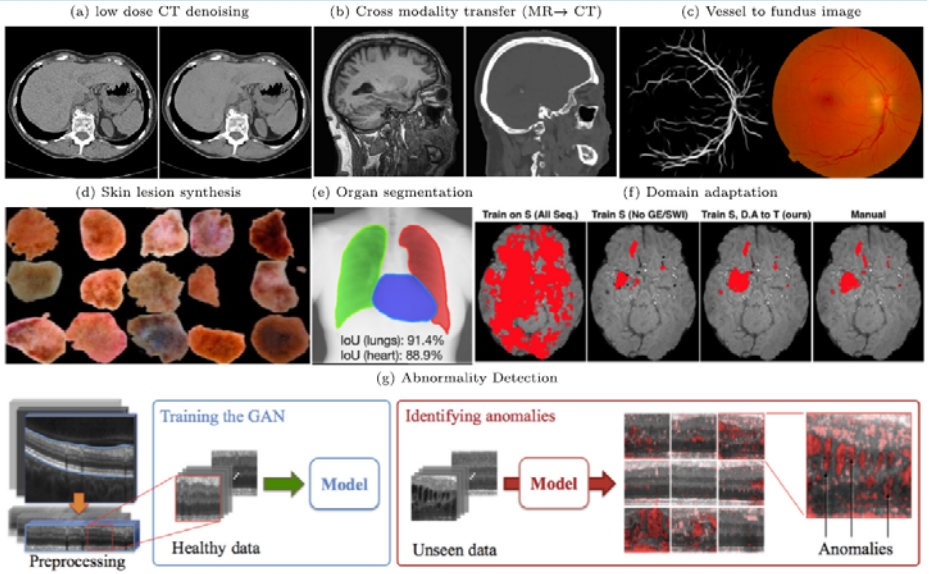
GAN related papers



Canonical tasks | Imaging modality | Number of GAN related papers

X. Yi, E. Walia and P. Babyn / Medical Image Analysis 58 (2019) 101552

Example applications using GANs



X. Yi, E. Walia and P. Babyn / Medical Image Analysis 58 (2019) 101552

Cross modality image synthesis

Pulications	Method	Remarks
MR → CT		
Nie et al. (2017, 2018)	Cascade GAN	[√]Brain; Pelvis
Emami et al. (2018)	cGAN	[√]Brain
CT → MR		
Jin et al. (2019)	CycleGAN	[✗] Brain
Jiang et al. (2018)	CycleGAN*	[✗] Lung
MR ↔ CT		
Chartsias et al. (2017)	CycleGAN	[✗] Heart
Zhang et al. (2018d)	CycleGAN*	[✗][3D] Heart
Huo et al. (2018)	CycleGAN*	[✗] Spleen
Chartsias et al. (2017)	CycleGAN	[✗] Heart
Hiasa et al. (2018)	CycleGAN*	[✗] Musculoskeletal
Wolterink et al. (2017a)	CycleGAN	[✗] Brain
Huo et al. (2018b)	CycleGAN	[✗] Abdomen
Yang et al. (2018b)	CycleGAN*	[✗] Brain
Maspero et al. (2018)	pix2pix	[√] Pelvis
CT → PET		
Bi et al. (2017)	cGAN	[√] Chest
Ben-Cohen et al. (2018)	FCN+cGAN	[√] Liver
PET → CT		
Armanious et al. (2018c)	cGAN*	[√] Brain
MR → PET		
Wei et al. (2018)	cascade cGAN	[√] Brain
Pan et al. (2018)	3D CycleGAN	[√] Brain
PET → MR		
Choi and Lee (2017)	pix2pix	[√] Brain
Synthetic → Real		
Hou et al. (2017)	synthesizer+cGAN	[√] Histopathology
Real → Synthetic		
Mahmood et al. (2018)	cGAN	[✗] Endocscopy
Zhang et al. (2018c)	CycleGAN*	[✗] X-ray
Domain adaption		
Chen et al. (2018a)	CycleGAN*	[✗] X-ray
T1 ↔ T2 MR		
Dar et al. (2019)	CycleGAN	[✗] Brain
Yang et al. (2018c)	cGAN	[✗] Brain
Welander et al. (2018)	CycleGAN, UNIT	[✗] Brain
Liu (2018)	CycleGAN	[✗] Knee
T1 → FLAIR MR		
Yu et al. (2018)	cGAN	[√] [3D] Brain
T1, T2 → MRA		
Olut et al. (2018)	pix2pix*	[√] Brain
3T → 7T MR		
Nie et al. (2018)	Cascade GAN	[√] Brain
Histopathology color normalization		
Bentaieb and Hamarneh (2018)	cGAN+classifier	[✗]
Zanjani et al. (2018)	InfoGAN	[✗]
Shaban et al. (2019)	CycleGAN	[✗]
Hyperspectral histology → H&E		
Bayramoglu et al. (2017a)	cGAN	[√] Lung

X. Yi, E. Walia and P. Babyn / Medical Image Analysis 58 (2019) 101552

Image-to-image translation

- 이미지 쌍으로 만들어진 데이터셋을 사용하여 input 이미지와 output 이미지를 매핑하는 것을 목표로 하는 생성모델의 한 분야
  - 예) 흑백 이미지 -> 컬러 이미지

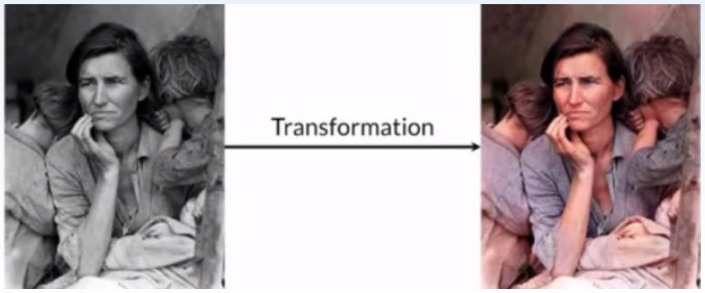
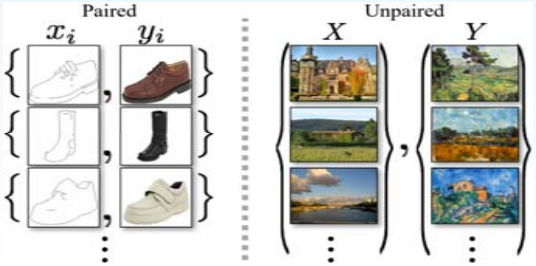


Image-to-image translation

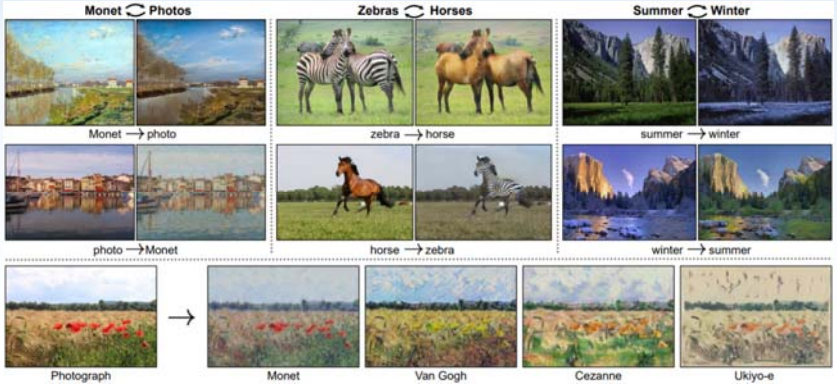
- **Problem statement**
  - 기존 학습 데이터 중 대부분의 경우 쌍을 이룬 학습 데이터 부족
  - 쌍을 이룬 학습 데이터 셋을 구하는 것은 어렵고 비용이 많이 듦



Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

Unpaired image-to-image translation

- **Unpaired image-to-image translation using cycle-consistent adversarial networks**



Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

Thank You!

AIML Lab.
https://sites.google.com/view/aiml-hallym
dongok.won@hallym.ac.kr