GlobalLogic®

Usage of Generative Adversarial Networks in Healthcare

Vladyslav Kolbasin

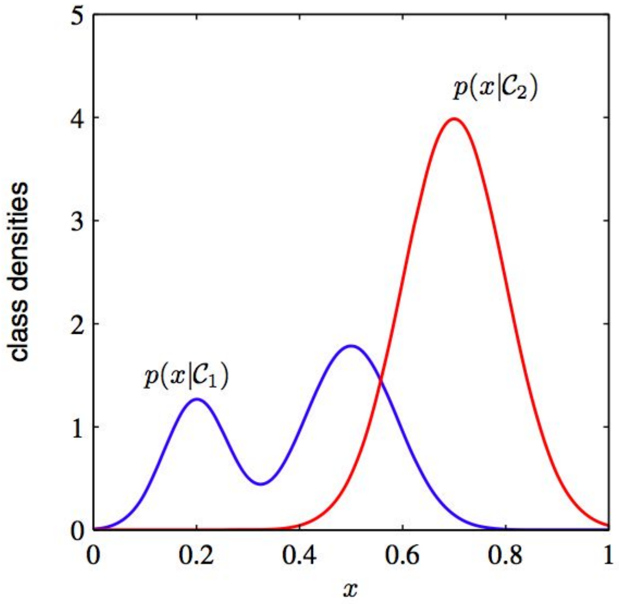
Lead Software Developer, Consultant

24.09.2017

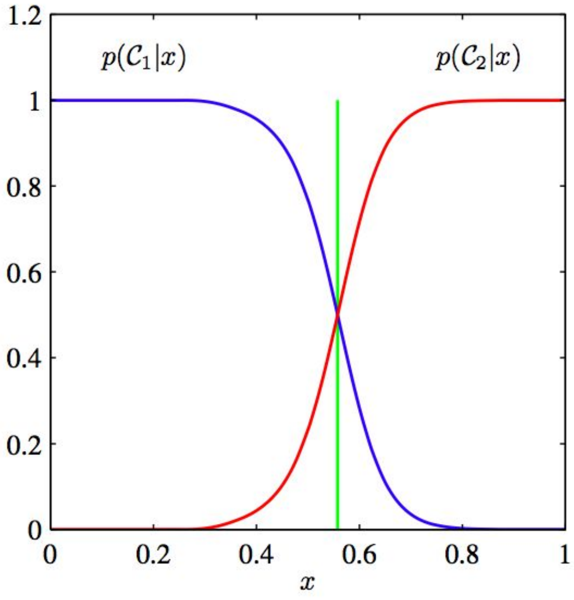
Agenda

1. Introduction to generative adversarial network
2. GAN theory. How to train it?
3. Several words about healthcare
4. How GANs are used in healthcare. Examples.
    1. Time-series data
    2. Image processing/generating/segmentation
5. Lesion segmentation
6. Conclusion

Generative *vs* discriminative models

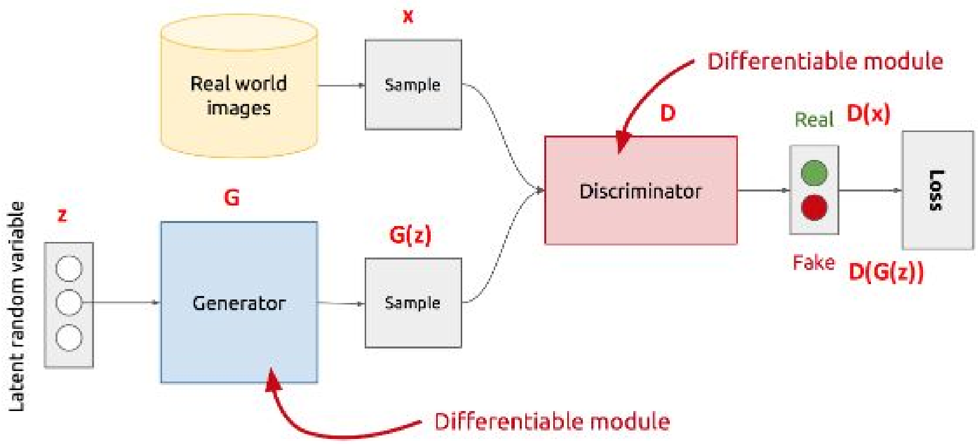


P(x|y)



P(y|x)

Generative adversarial network. Architecture



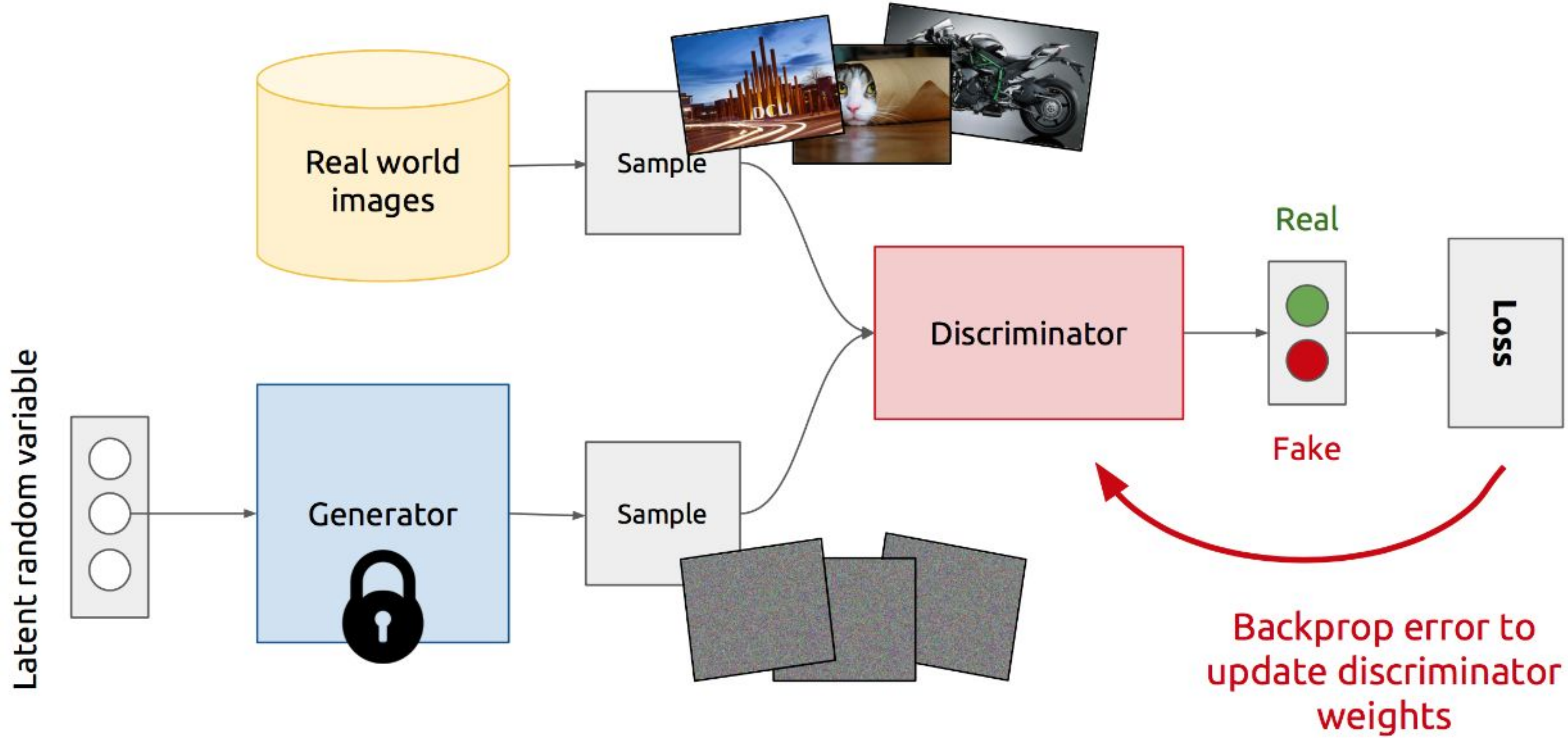
- GAN – two neural networks competing against each other in a zero-sum game framework. (Ian Goodfellow et al. in 2014)
- G tries to “trick” D by generating samples that are hard for D to distinguish from data
- Some kind of unsupervised learning

- Networks try to:
  - D(G(z)) => max
  - D(x)(1 — D(G(z))) => max

  => Nash equilibrium: $(\boldsymbol{\theta}^{(D)}, \boldsymbol{\theta}^{(G)})$

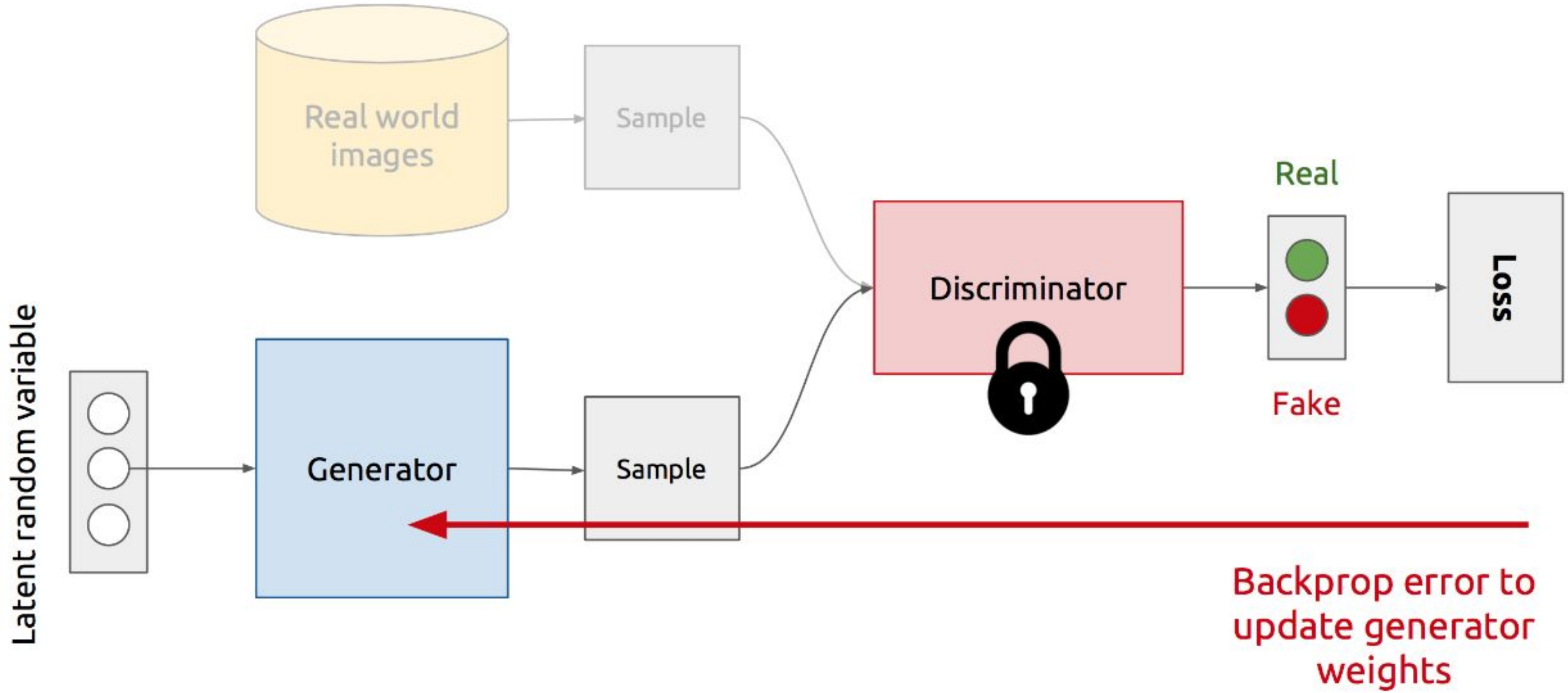
Credit:
https://www.slideshare.net/xavigiro/deep-learning-for-computer-vision-generative-models-and-adversarial-training-upc-2016

GAN. Training Discriminator



Credit: https://www.slideshare.net/xavigiro/deep-learning-for-computer-vision-generative-models-and-adversarial-training-upc-2016

GAN. Training Generator



Credit: https://www.slideshare.net/xavigiro/deep-learning-for-computer-vision-generative-models-and-adversarial-training-upc-2016

GAN. Training algorithm

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k, is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

  for k steps **do**

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
  - Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right].$$

  end for

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

GAN. How to train?

- Optimization criteria:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

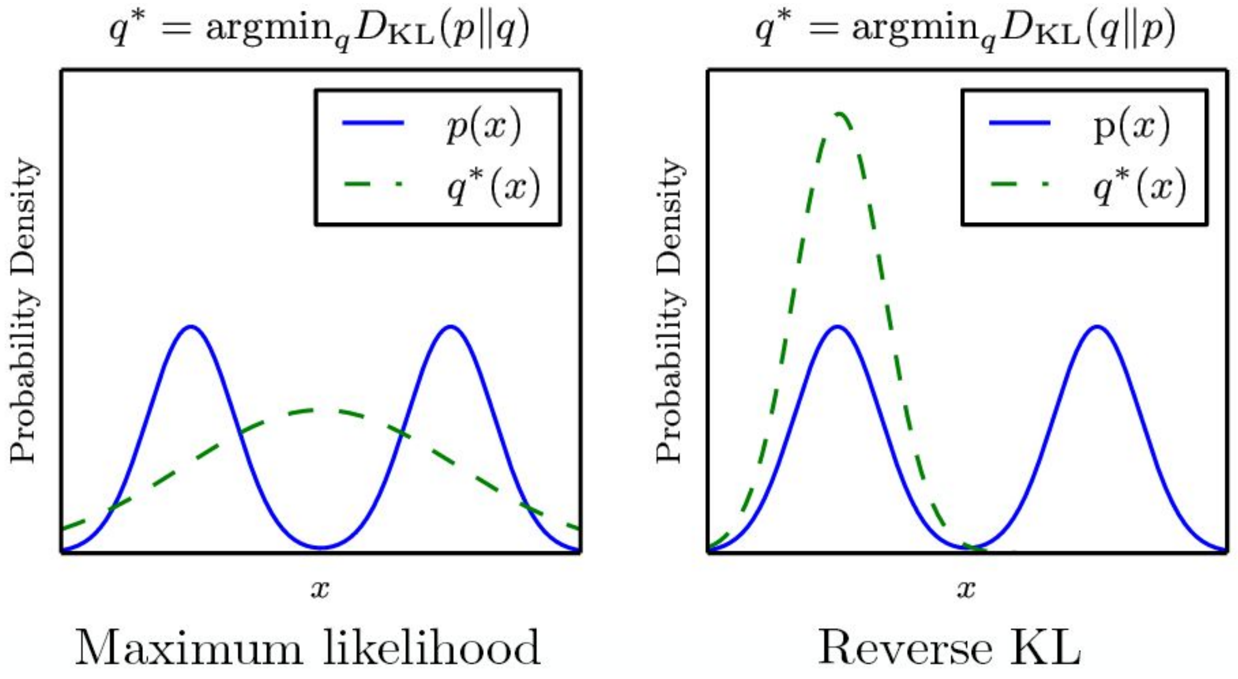
True data

Noise for generating data

- We want to make distributions equal
  - It is equivalent to maximizing log-likelihood or
  - KL-divergence: $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} D_{\text{KL}}\left(p_{\text{data}}(\boldsymbol{x}) \| p_{\text{model}}(\boldsymbol{x};\boldsymbol{\theta})\right)$
  - We can use other divergence too
    - Jensen-Shannon divergence $D_{JS}(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M), M = \frac{1}{2}(P+Q)$

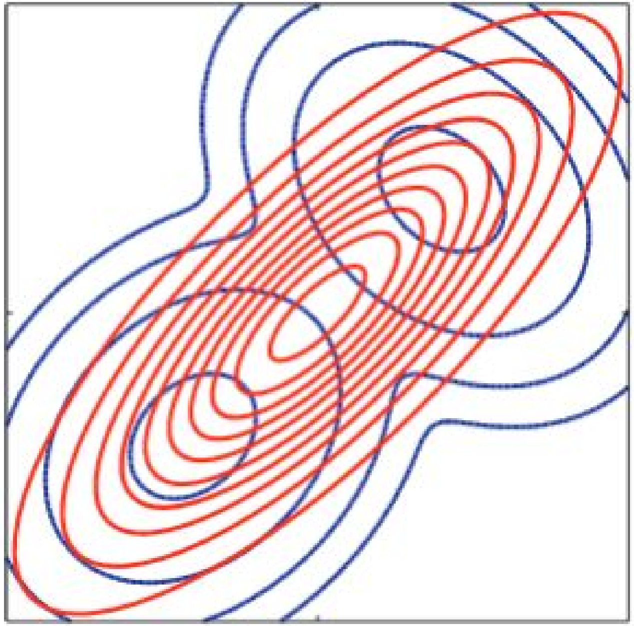
KL-divergence

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i)\,\log \frac{P(i)}{Q(i)}$$

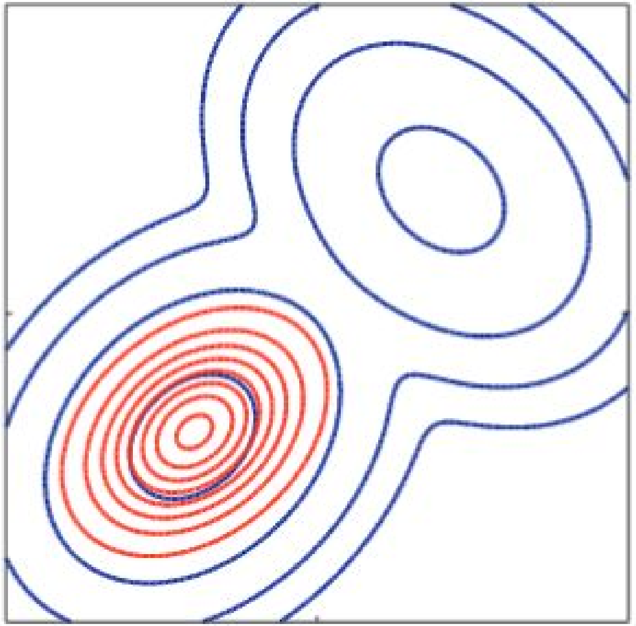


KL-divergence

$$q^* = \text{argmin}_q D_{\text{KL}}(p||q)$$



$$q^* = \text{argmin}_q D_{\text{KL}}(q||p)$$



Demo. Mode-collapse. Oscillations in GAN

Fixing mode collapse. wGAN(s)

Wasserstein GAN. https://arxiv.org/pdf/1701.07875.pdf

- Introduced other metric: Wasserstein metric (earth mover's distance)

$$W(P_r, P_\theta) = \inf_{\gamma \in \Pi(P_r, P_\theta)} \mathbb{E}_{(x,y)\sim\gamma}[\| x - y \|]$$

- Summary about wGAN:
    - **For mathematicians**: it uses Wasserstein distance instead of Jensen-Shannon divergence to compare distributions
    - **For engineers**: it gets rid of a few unnecessary logarithms, and clips weights
    - **For others**: it employs an art critic instead of forgery expert

- More math details: https://www.cph-ai-lab.com/wasserstein-gan-wgan

Fixing mode collapse. Improved wGAN.

Improved training of Wasserstein GAN. https://arxiv.org/pdf/1704.00028.pdf

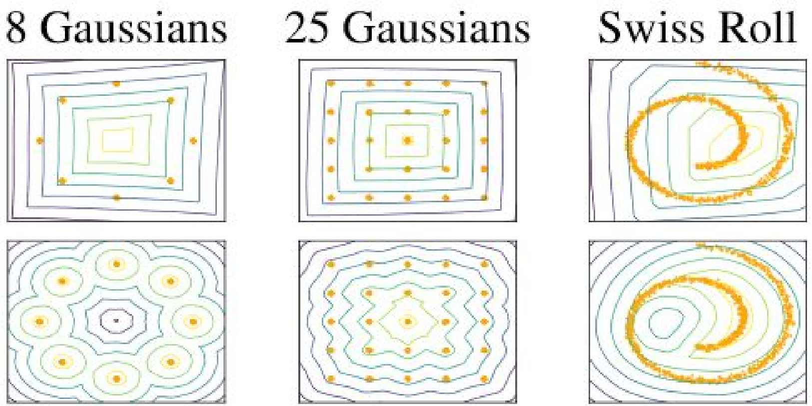
- Penalize the norm of the gradient of the critic with respect to its input instead of clipping weights.
- This 'gradient penalty' is simply added to the Wasserstein distance for the total loss.

Algorithm 1 WGAN with gradient penalty. We use default values of $\lambda = 10$, $n_{\text{critic}} = 5$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$.

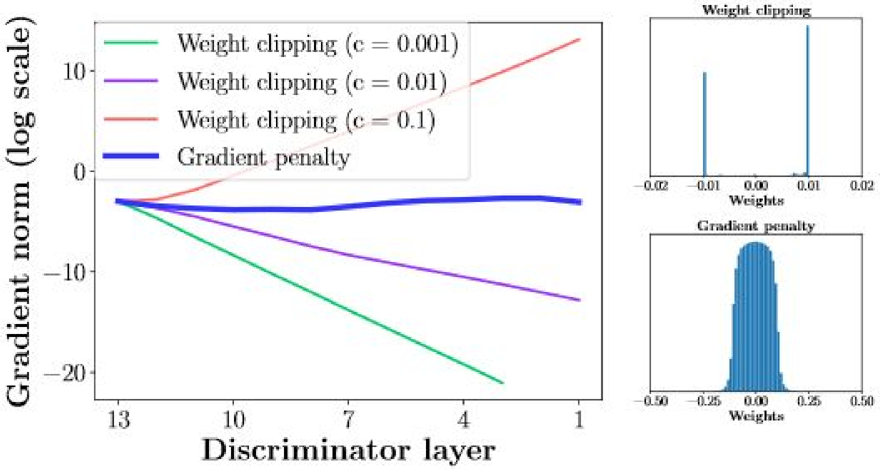
Require: The gradient penalty coefficient λ, the number of critic iterations per generator iteration n_{critic}, the batch size m, Adam hyperparameters α, β_1, β_2.
Require: initial critic parameters w_0, initial generator parameters θ_0.

while θ has not converged **do**
  for $t = 1, \ldots, n_{\text{critic}}$ **do**
    for $i = 1, \ldots, m$ **do**
      Sample real data $\boldsymbol{x} \sim \mathbb{P}_r$, latent variable $\boldsymbol{z} \sim p(\boldsymbol{z})$, a random number $\epsilon \sim U[0,1]$.
      $\tilde{\boldsymbol{x}} \leftarrow G_\theta(\boldsymbol{z})$
      $\hat{\boldsymbol{x}} \leftarrow \epsilon \boldsymbol{x} + (1-\epsilon)\tilde{\boldsymbol{x}}$
      $L^{(i)} \leftarrow D_w(\tilde{\boldsymbol{x}}) - D_w(\boldsymbol{x}) + \lambda(\|\nabla_{\hat{\boldsymbol{x}}} D_w(\hat{\boldsymbol{x}})\|_2 - 1)^2$
    end for
    $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m}\sum_{i=1}^m L^{(i)}, w, \alpha, \beta_1, \beta_2)$
  end for
  Sample a batch of latent variables $\{\boldsymbol{z}^{(i)}\}_{i=1}^m \sim p(\boldsymbol{z})$.
  $\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m}\sum_{i=1}^m -D_w(G_\theta(\boldsymbol{z})), \theta, \alpha, \beta_1, \beta_2)$
end while

Improved WGAN: +gradient penalty



(a) Value surfaces of WGAN critics trained to optimality on toy datasets using (top) weight clipping and (bottom) gradient penalty. Critics trained with weight clipping fail to capture higher moments of the data distribution. The 'generator' is held fixed at the real data plus Gaussian noise.

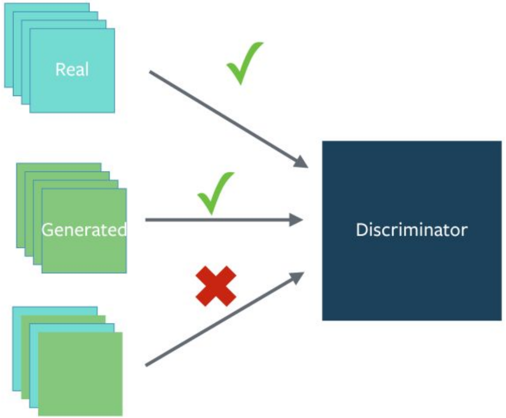


(b) (left) Gradient norms of deep WGAN critics during training on toy datasets either explode or vanish when using weight clipping, but not when using a gradient penalty. (right) Weight clipping (top) pushes weights towards two values (the extremes of the clipping range), unlike gradient penalty (bottom).

Demo. Mode-collapse.
Fixing oscillations in GANs

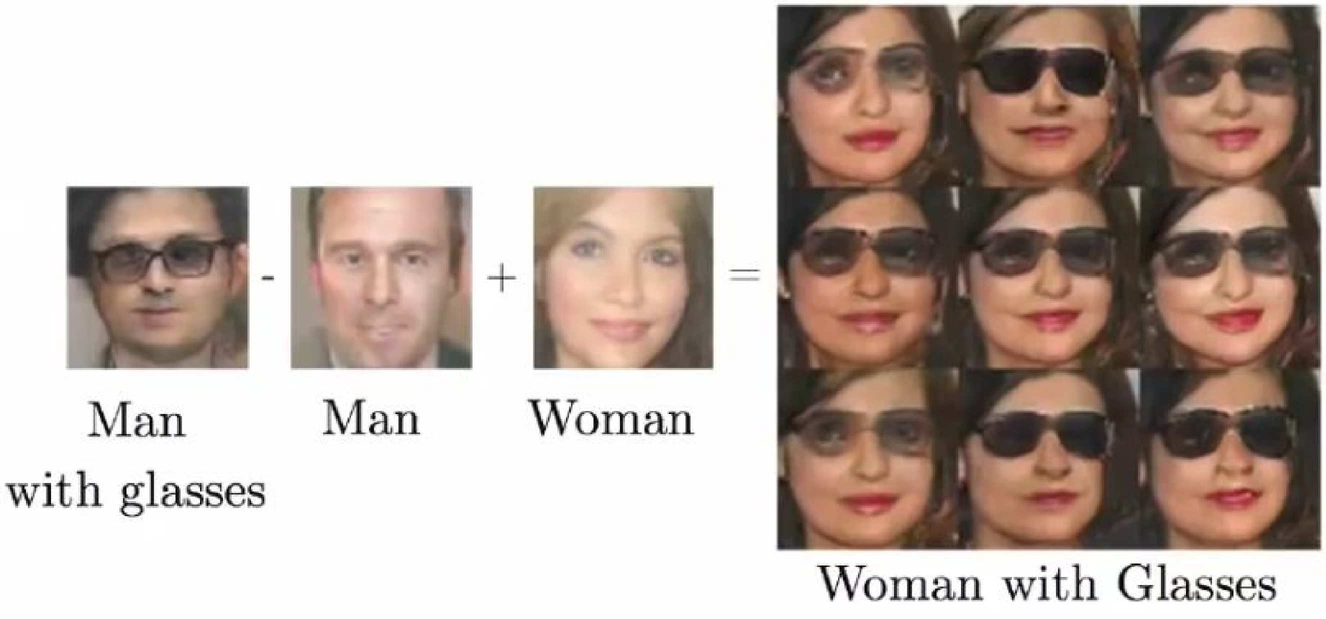
GAN hacks

- https://github.com/soumith/ganhacks
- About 17 hacks:
    - Normalize the inputs
    - Use Soft and Noisy Labels
    - Avoid Sparse Gradients: ReLU, MaxPool
    - Use SGD for discriminator and ADAM for generator
    - …
    - Batch Normalization

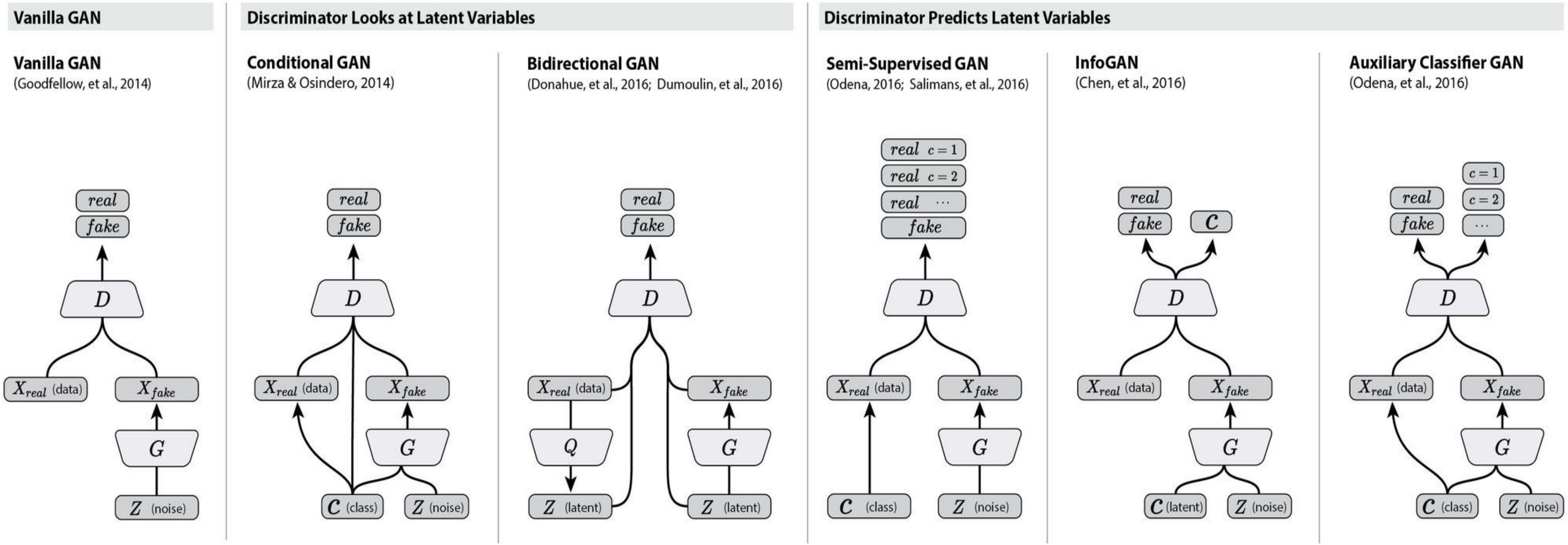


Vector Space Arithmetics

- “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Models”. https://arxiv.org/pdf/1511.06434.pdf



GANs Taxonomy



Healthcare

Deep Learning in Healthcare, Pros

- Deep learning can provide huge positive impact in Healthcare

Google Research DeepMind Health

- Healthcare is one of the most popular area in DL researches
    - List of significant deep learning papers in healthcare: https://github.com/albarqouni/Deep-Learning-for-Medical-Applications
- Main ideas:
    - Improve the health of populations
    - Lower the cost of care
    - Improve the patient experience

DL in Healthcare, Tasks

- Mining Medical Data
  - There is huge amount of doctors' reports, medical images and test results
- Better or earlier diagnostics
  - Early diagnostics without complex hardware (for example risk of heart attack)
  - Predict next visit to a doctor
  - Automatic MRI slides segmentation and so on
- Better, faster diagnoses
- More effective drug treatment
- Genomics for personalized medicine
- Creating new drugs
- And so on…

Research issues with healthcare

- Health data is highly regulated
  - USA: FDA + HIPAA (http://ww.fda.gov)
  - EU: Directives 95/46/EC, 98/79/EC and others (http://ec.europa.eu/health/medical-devices/index_en.htm)

- Data labelling is challenging
  - There are only tens of thousands people **trained** and **licensed** to annotate all of those observable features

- Health data is highly multimodal
  - There are **dozens** of kinds of medical imaging devices each producing images according to their respective physical principles

Health Data

Definitions

- Identifiable Health Data – any information about health status, provision of health care, or payment for health care that can be linked to an individual
- Privacy of Health Data allows a person to keep their medical records from being revealed to others

The Main Challenge

- The challenge in data privacy is to share data while protecting personally identifiable information

USA & EU Regulations

- There are different requirements to data protection depending on what country citizens the data belong to

What is HIPAA?

Abbreviation

- **H**ealth
- **I**nsurance
- **P**ortability and
- **A**ccountability
- **A**ct of 1996

Purpose

- Congress passed the federal law to provide consumers with greater access to health care insurance, to protect the privacy of health care data, and to promote more standardization and efficiency in the healthcare industry.
- HIPAA was designed to allow portability of health insurance between jobs.
- HIPAA directed the Department of Health and Human Services (DHHS) to issue federal regulations. DHHS has issued HIPAA privacy regulations as well as other regulations under HIPAA.

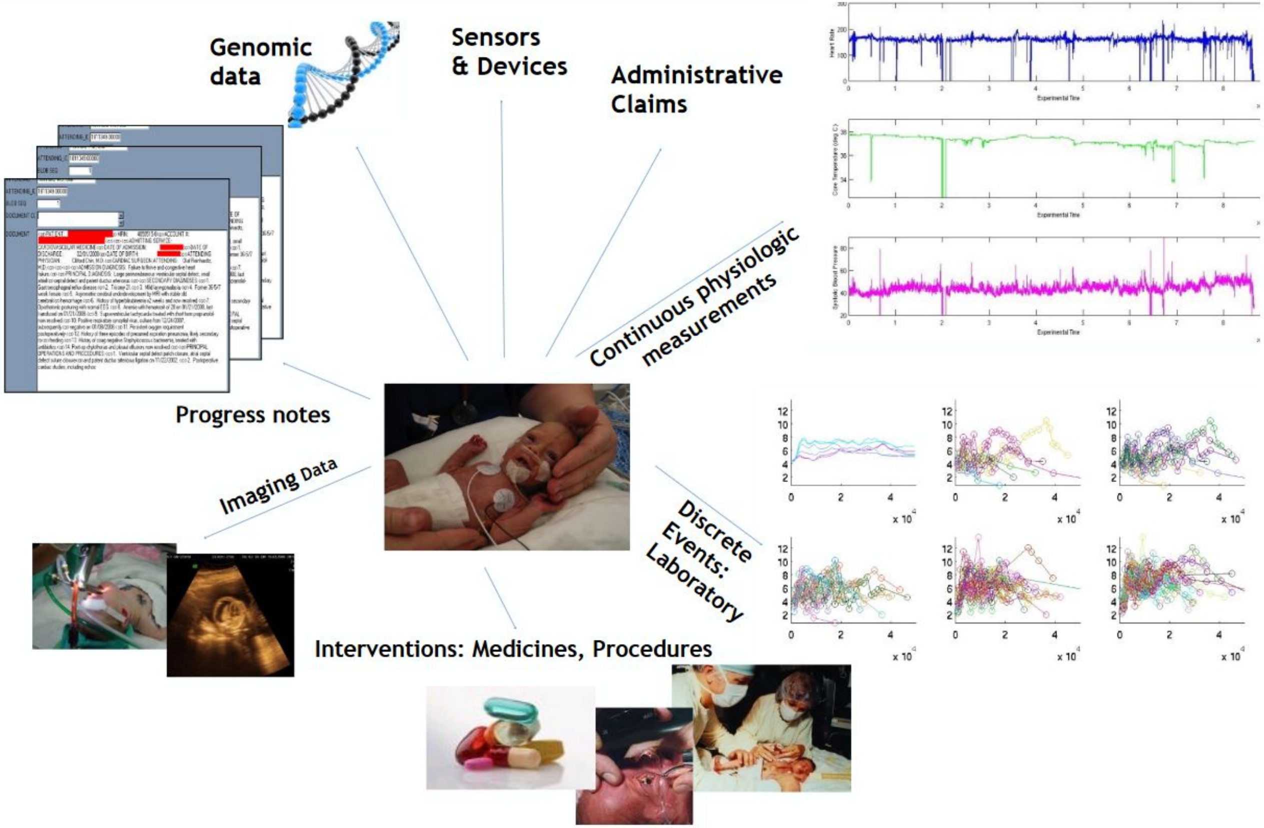
Applications

- Health Data Generation
- Medical Image Segmentation
- Anomaly Detection

1. Health data generation

- Types of data:
  - Electronic Health Records (EHR)
  - Medical Images
  - Other data...
- Accessing to EHR is highly regulated (restricted)
- Need to do de-identification of EHR data
  - HIPAA is watching you….

Electronic Health Records



1.1 Health data generation. EHR

- “Generating Multi-label Discrete Patient Records using Generative Adversarial Network”. https://arxiv.org/pdf/1703.06490.pdf

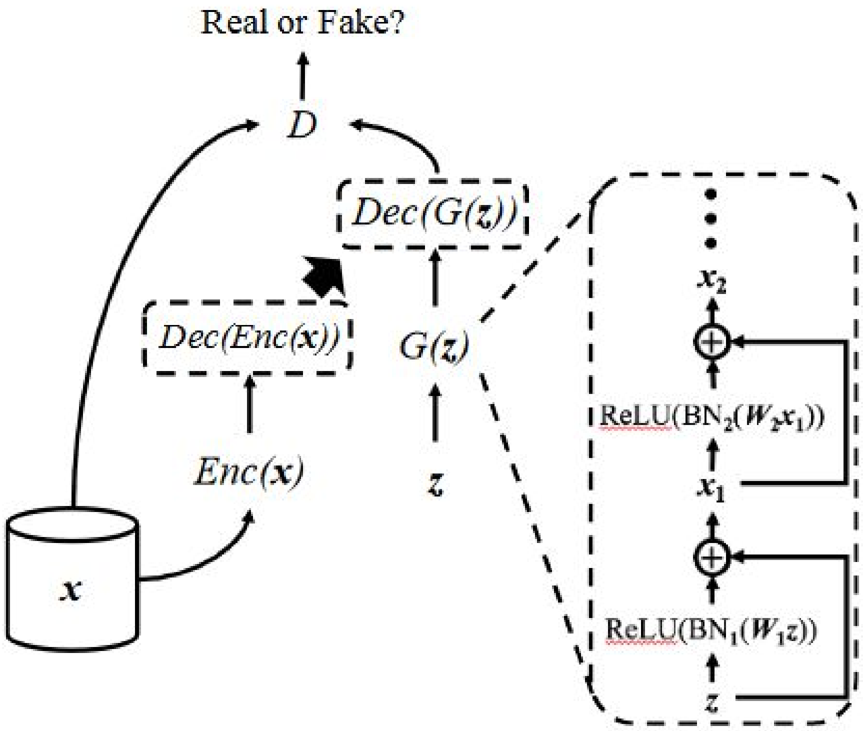
- Medical Generative Adversarial Network (medGAN) generates realistic synthetic patient records
    - High-dimensional discrete variables

Table 1: Basic statistics of datasets A, B and C

Dataset	A	B	C
# of patients	258,559	46,520	30,738
# of unique codes	615	1071	569
Avg. # of codes per patient	38.37	11.27	53.02
Max # of codes for a patient	198	90	871
Min # of codes for a patient	1	1	2

medGAN. Architecture

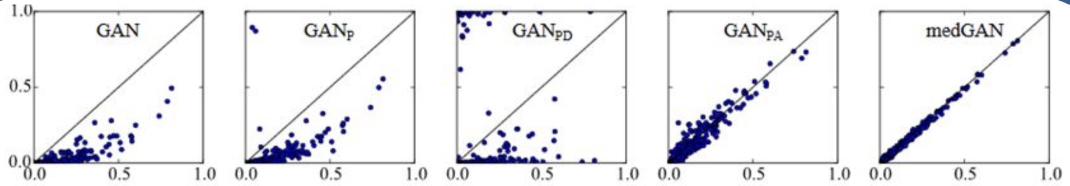
Figure 1: Architecture of medGAN: The discrete $\mathbf{x}$ comes from the source EHR data, $\mathbf{z}$ is the random prior for the generator G; G is a feedforward network with shortcut connections (righthand side figure); An autoencoder (i.e, the encoder Enc and decoder Dec) is learned from $\mathbf{x}$; The same decoder Dec is used after the generator G to construct the discrete output. The discriminator D tries to differentiate real input $\mathbf{x}$ and discrete synthetic output $Dec(G(\mathbf{z}))$.



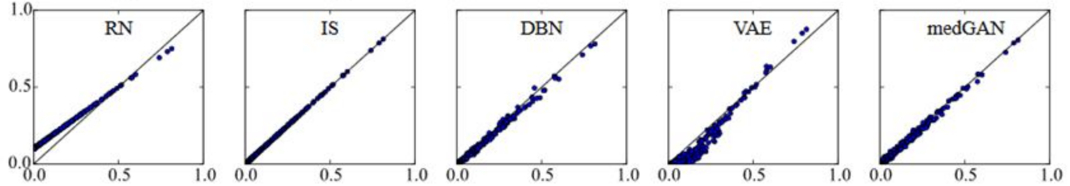
medGAN. Models for comparison

- **GAN**: the same architecture as medGAN with the standard training strategy, but do not pre-train the autoencoder.
- **GAN-P**: pre-train the autoencoder (in addition to the GAN).
- **GAN-PD**: pre-train the autoencoder and use minibatch discrimination.
- **GAN-PA**: pre-train the autoencoder and use minibatch averaging.
- **medGAN**: pre-train the autoencoder and use minibatch averaging.

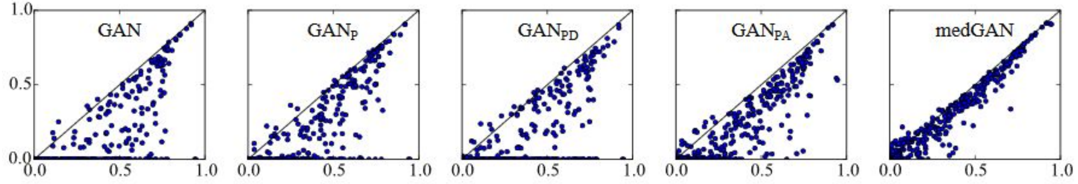
Also use batch normalization and a shortcut connection for the generator G.



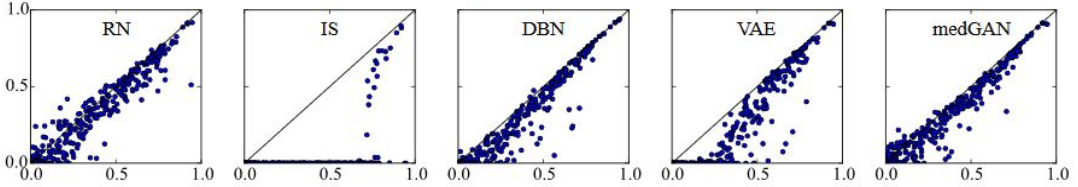
(a) Dimension-wise probability performance of various versions of medGAN.



(b) Dimension-wise probability performance of baseline models and medGAN.



(a) Dimension-wise prediction performance of various versions of medGAN.



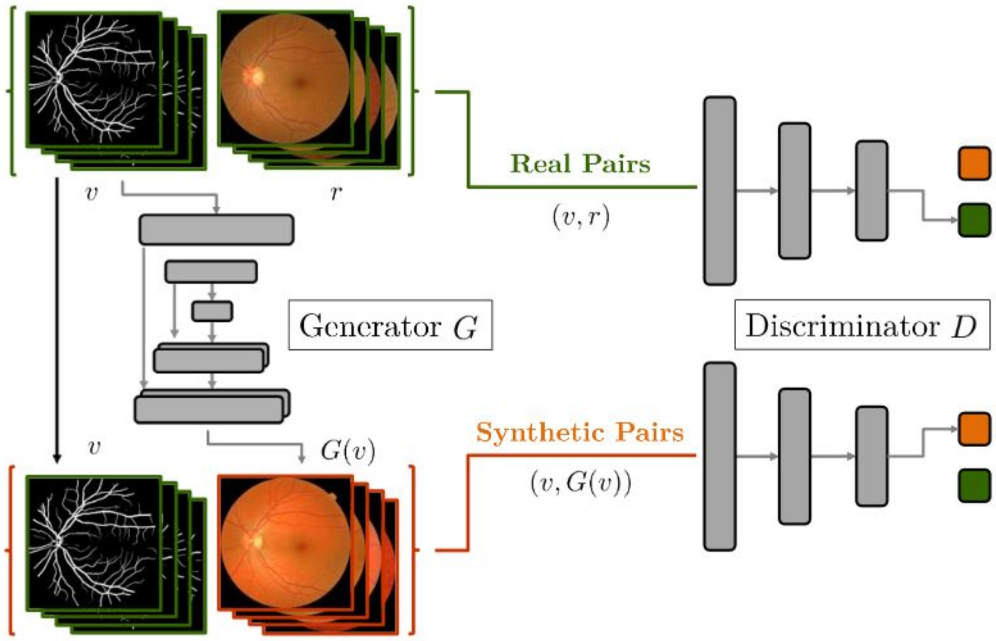
(b) Dimension-wise prediction performance of baseline models and medGAN.

Scatterplots of dimension-wise probability results. Each dot represents one of 615 codes. The diagonal line indicates the ideal performance where the real and synthetic data show identical quality

1) The x-axis represents the Bernoulli success probability for the real dataset A, and y-axis the probability for the synthetic counterpart generated by each model.

2) The x-axis represents the F1-score of the logistic regression classifier trained on the real dataset A. The y-axis represents the F1-score of the classifier trained on the synthetic counterpart generated by each model.

1.2. Image-to-Image translation using GANs



- “Towards Adversarial Retinal Image Synthesis” https://arxiv.org/pdf/1701.08974v1.pdf
- Mapping from a binary vessel tree to a new retinal image

- Aims:
  - Validating image analysis techniques
  - Medical training
  - Therapy planning

- Used ideas from “Image-to-Image Translation with Conditional Adversarial Networks” https://arxiv.org/pdf/1701.08974v1.pdf https://phillipi.github.io/pix2pix/

Retinal GAN. Loss modification

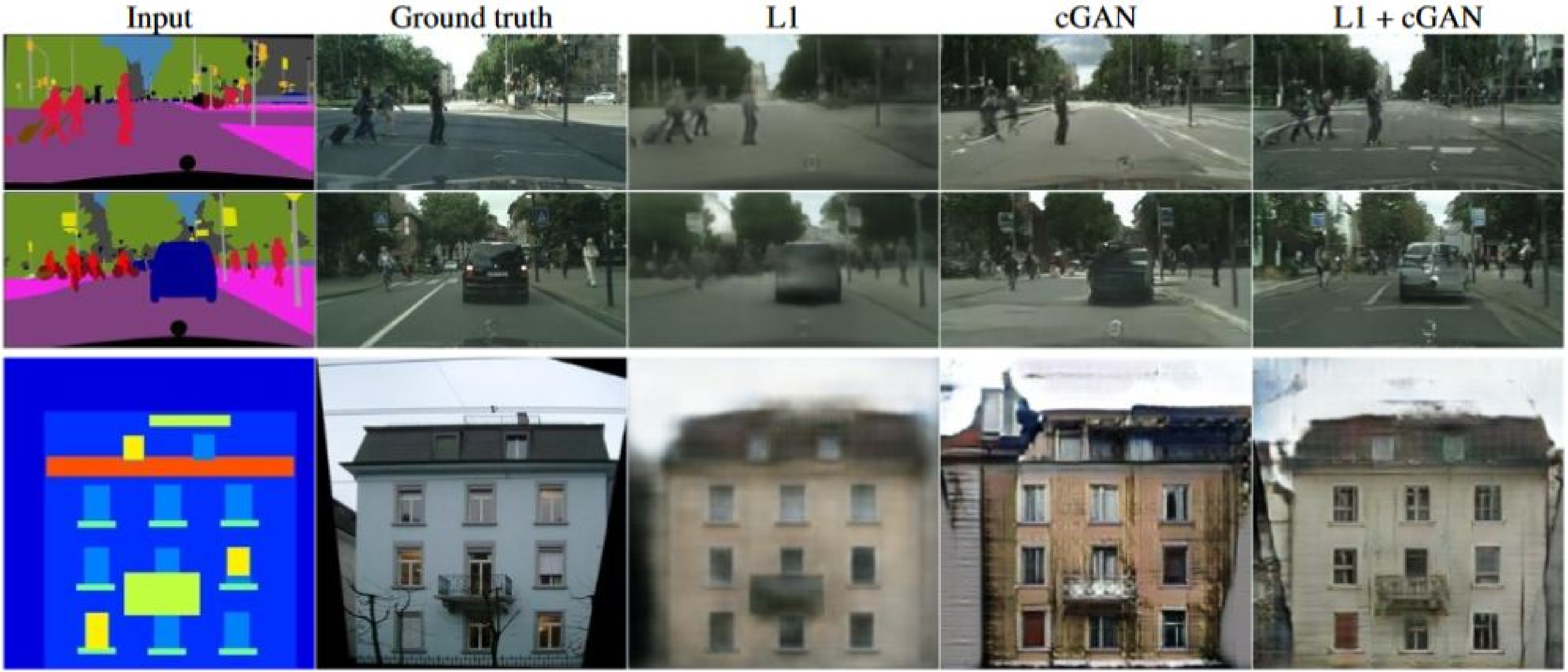
- GAN with usual adversarial loss will not produce realistic images

- Modified loss: combine the adversarial loss with a global L1 loss:

$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \lambda \mathbb{E}_{v,r \sim p_{data}(v,r)} \left(||r - G(v)||_1 \right)$$

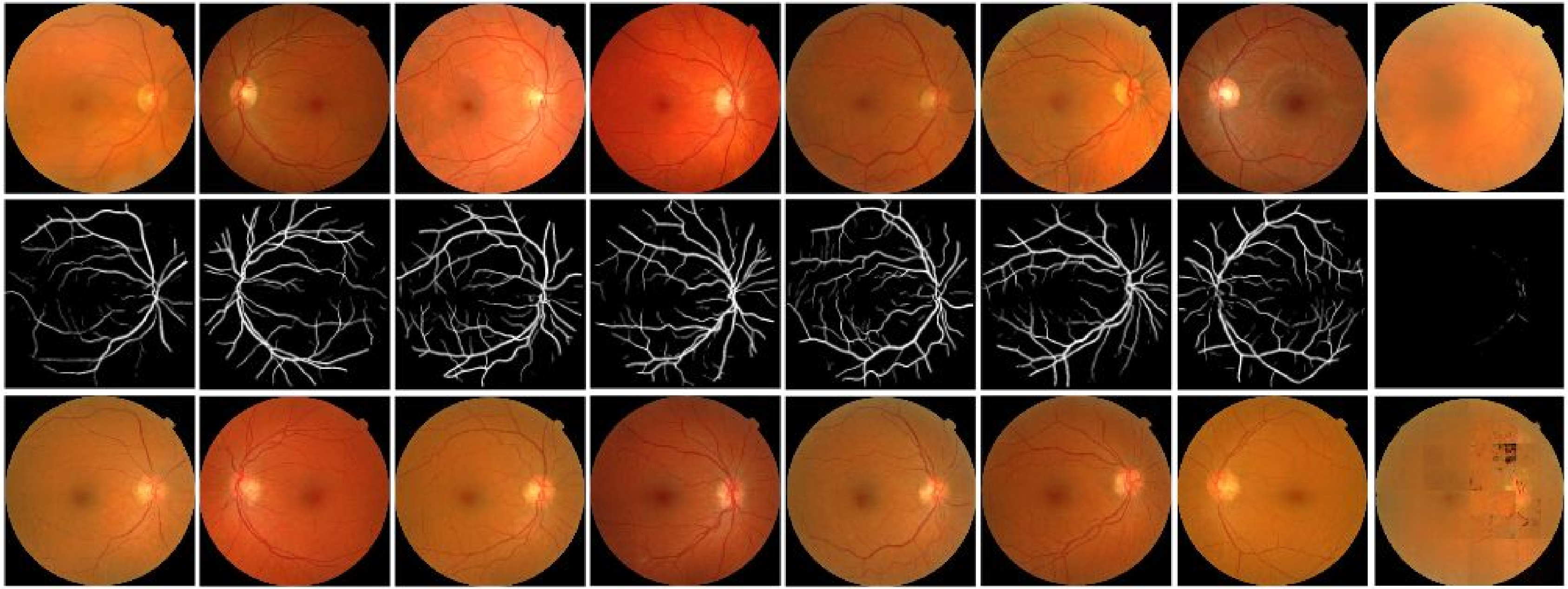
  - The L1 loss controls low-frequency information in images generated by G in order to produce globally consistent results
  - Adversarial loss promotes sharp results

Loss modification. Examples



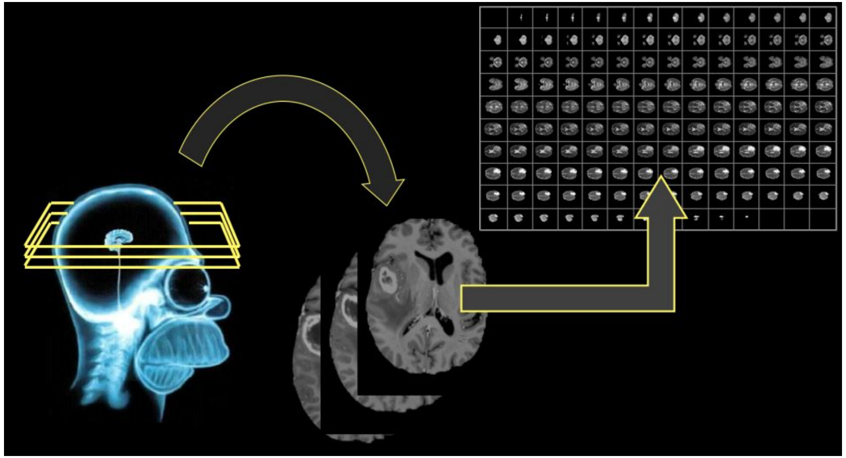
Credit: “Image-to-Image Translation with Conditional Adversarial Networks” https://arxiv.org/pdf/1701.08974v1.pdf

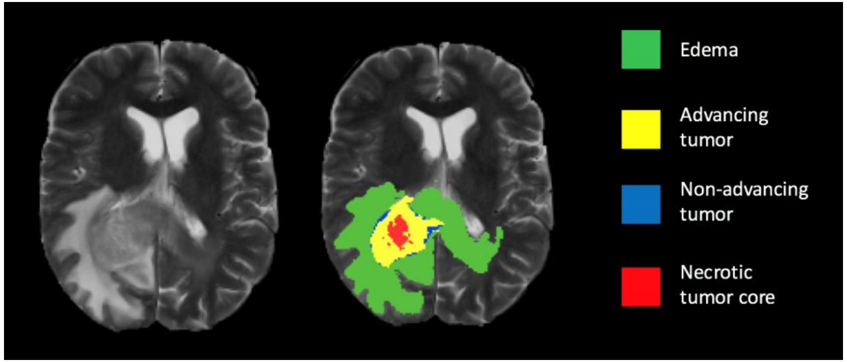
Retinal GAN. Results. Evaluation



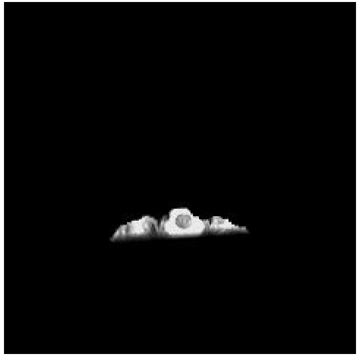
- Evaluation done with retinal image quality metrics which have been employed previously to assess the quality of retinal images

2. Medical image segmentation

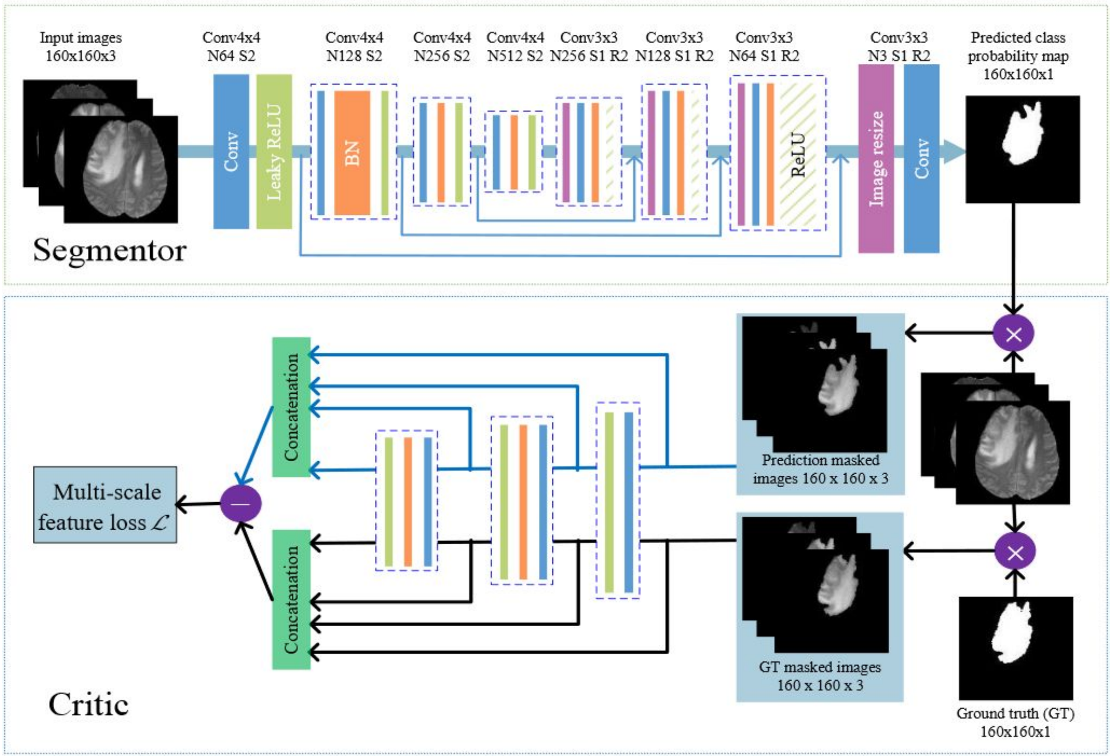




- **Brain Tumor Segmentation Challenge** http://www.braintumorsegmentation.org/
- Each subject in BRATS 2015 dataset is a 3D brain MRI volume with size 240×240×155



2. Medical image segmentation. SegAN



- “SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation” https://arxiv.org/pdf/1706.01805.pdf
- Semantic segmentation of images

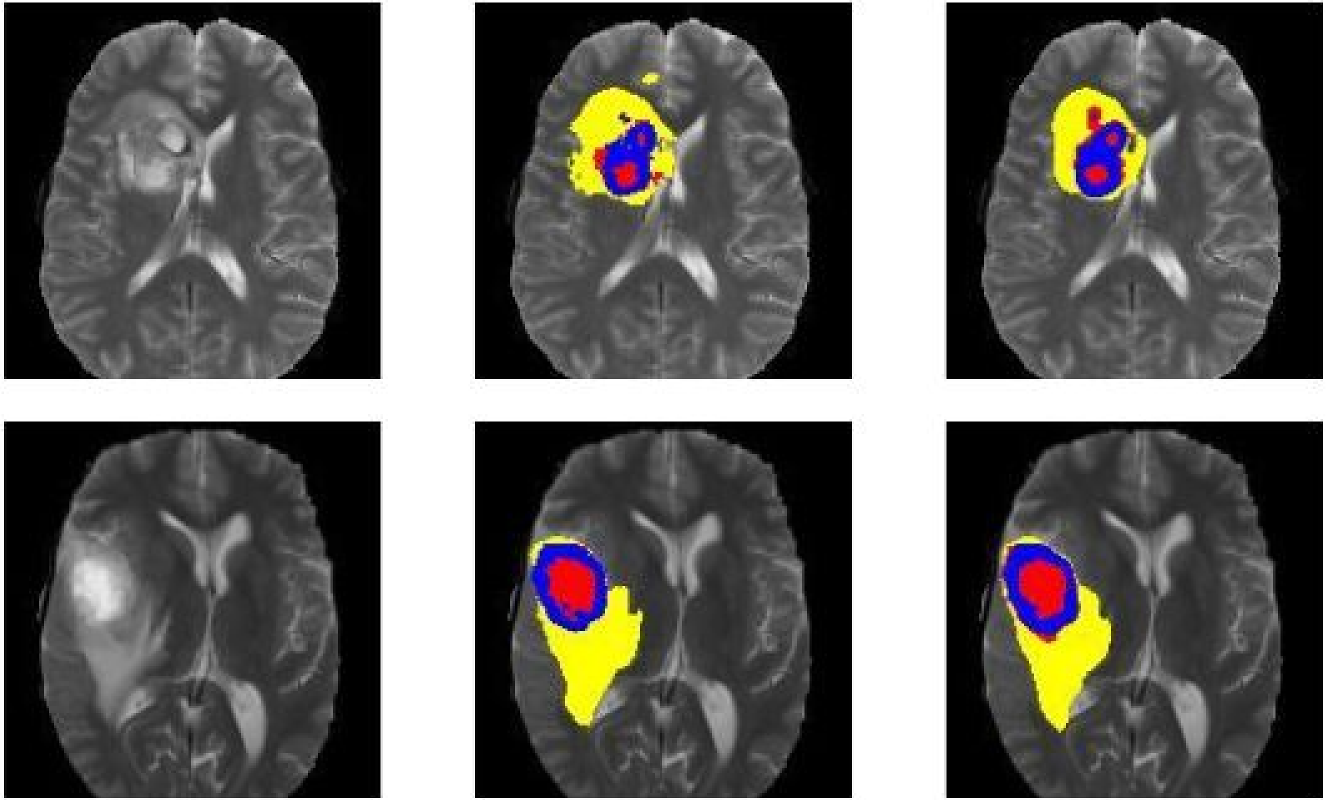
SegAN. Main ideas

- Usual single scalar output of Discriminator doesn’t provide sufficient gradient feedback to the networks
- Instead was used **multi-scale feature loss** to get more stable results:

$$\min_{\theta_S} \max_{\theta_C} \mathcal{L}(\theta_S, \theta_C) = \frac{1}{N} \sum_{n=1}^{N} \ell_{\mathrm{mae}}(f_C(x_n \circ S(x_n)), f_C(x_n \circ y_n))$$

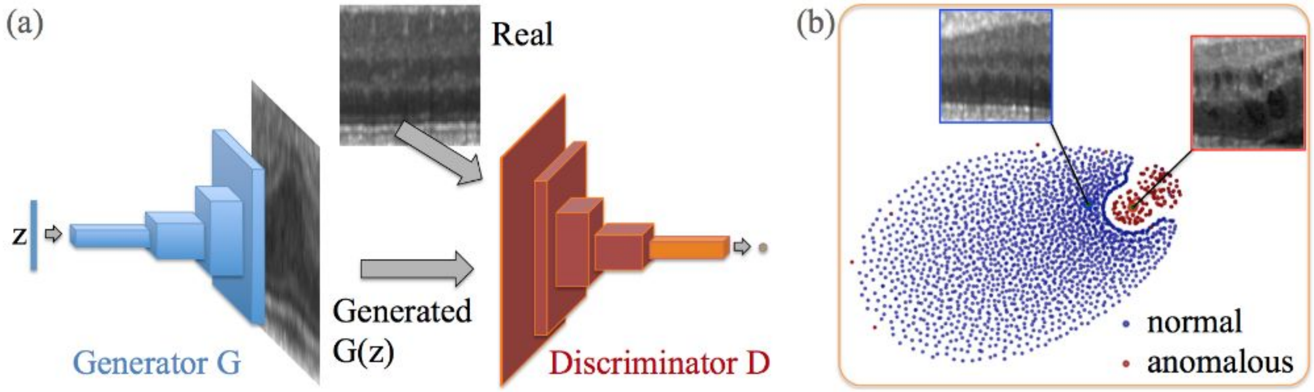
$$\ell_{\mathrm{mae}}(f_C(x), f_C(x')) = \frac{1}{L} \sum_{i=1}^{L} ||f_C^i(x) - f_C^i(x')||_1$$

Medical image segmentation. SegAN. Results



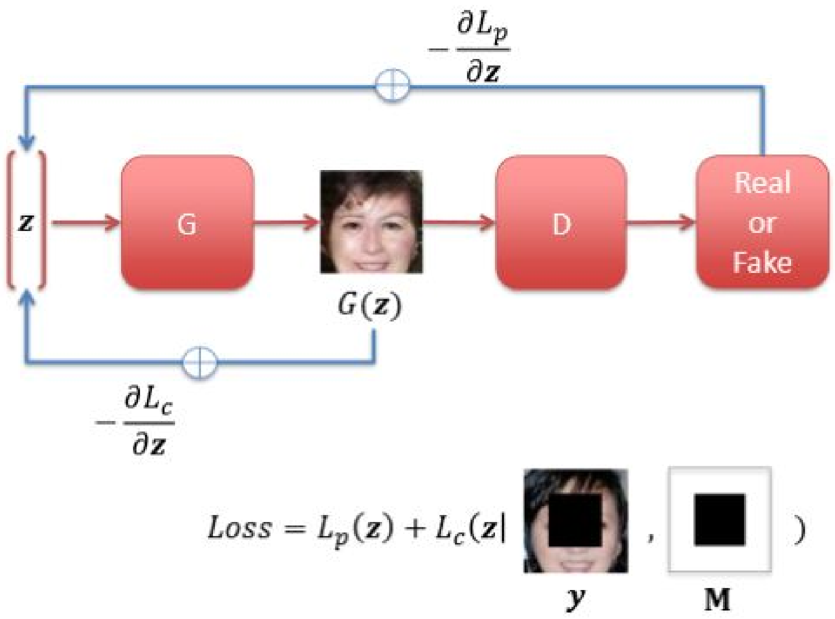
3. Anomaly detection

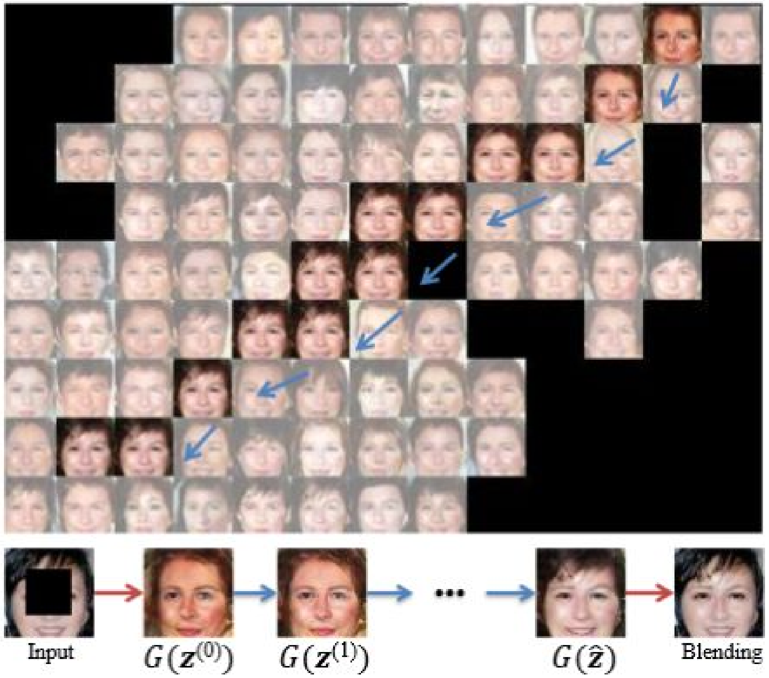
- “Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery”. https://arxiv.org/pdf/1703.05921.pdf



What about inverse mapping: image-latent space?

- “Semantic Image Inpainting with Deep Generative Models” https://arxiv.org/pdf/1607.07539.pdf





Mapping new images to the latent space

- Only the coefficients of **z** are adapted via backpropagation. The trained parameters of the generator and discriminator are kept fixed.
- Overall loss or **Anomaly score**:

$$\mathcal{L}(\mathbf{z}_\gamma) = (1-\lambda) \cdot \mathcal{L}_R(\mathbf{z}_\gamma) + \lambda \cdot \mathcal{L}_D(\mathbf{z}_\gamma)$$

- Anomaly score consists of two parts:
  - **Residual Loss** - visual similarity $\mathcal{L}_R(\mathbf{z}_\gamma) = \sum |\mathbf{x} - G(\mathbf{z}_\gamma)|$
  - **Discrimination Loss** - enforces the generated image to lie on the manifold

Improved discrimination loss based on feature matching

- f(.) – output of intermediate layer of the **discriminator**
  - It is some statistics of an input image

$$\mathcal{L}_D(\mathbf{z}_\gamma) = \sum |\mathbf{f}(\mathbf{x}) - \mathbf{f}(G(\mathbf{z}_\gamma))|$$

- This approach utilizes the trained discriminator not as classifier but as a **feature extractor**

Results for anomaly detection

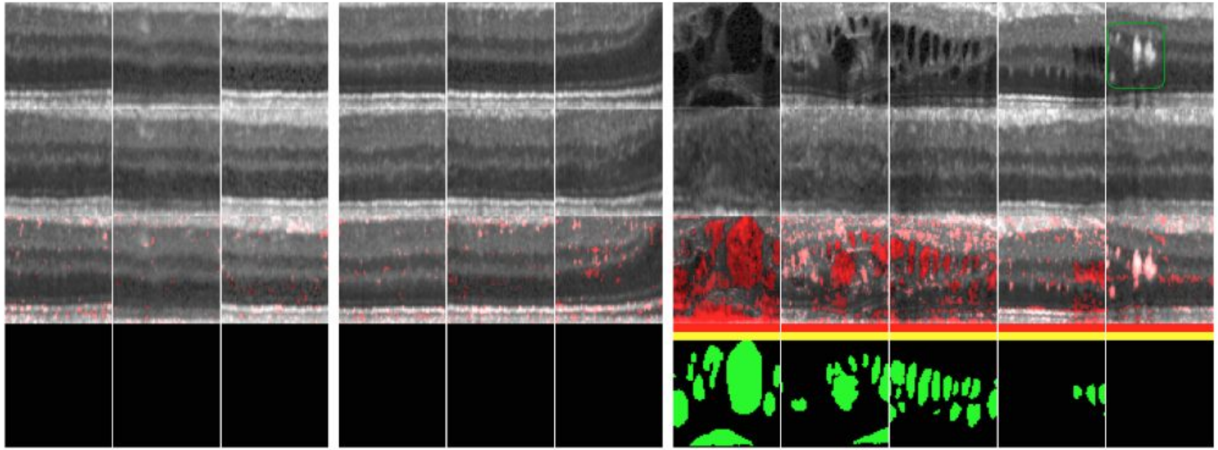
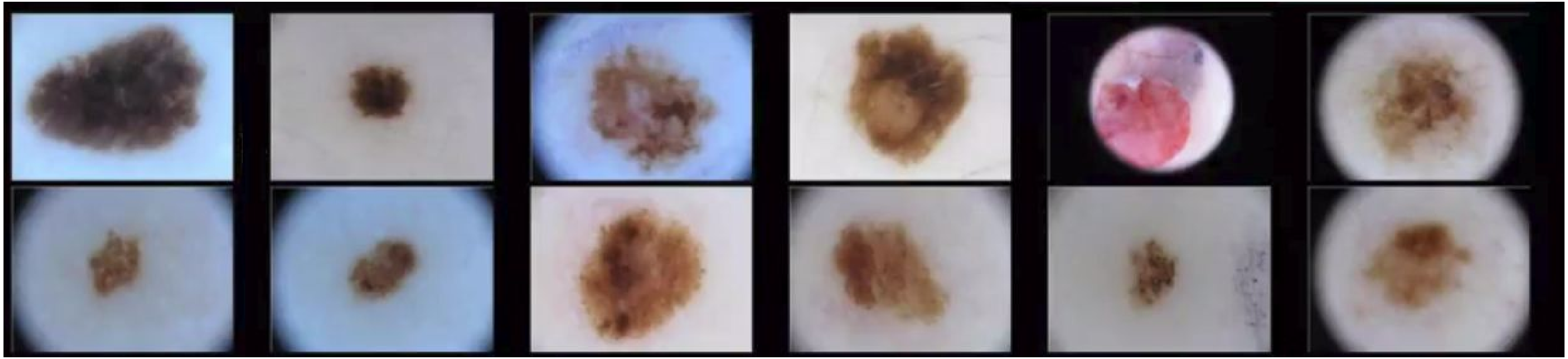


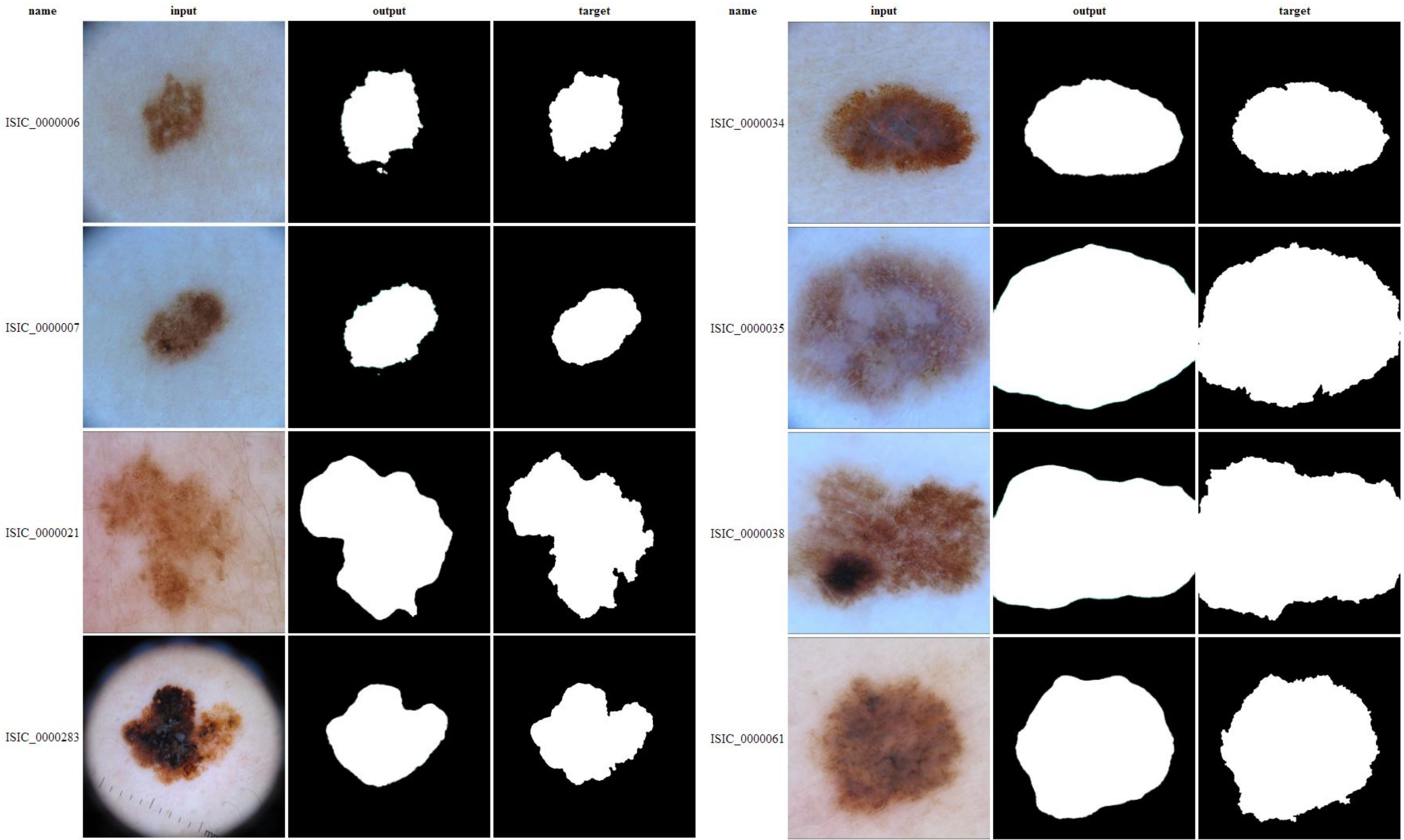
Fig. 3. Pixel-level identification of anomalies on exemplary images. First row: Real input images. Second row: Corresponding images generated by the model triggered by our proposed mapping approach. Third row: Residual overlay. Red bar: Anomaly identification by *residual score*. Yellow bar: Anomaly identification by *discrimination score*. Bottom row: Pixel-level annotations of retinal fluid. First block and second block: Normal images extracted from OCT volumes of healthy cases in the training set and test set, respectively. Third block: Images extracted from diseased cases in the test set. Last column: Hyperreflective foci (within green box). (Best viewed in color)

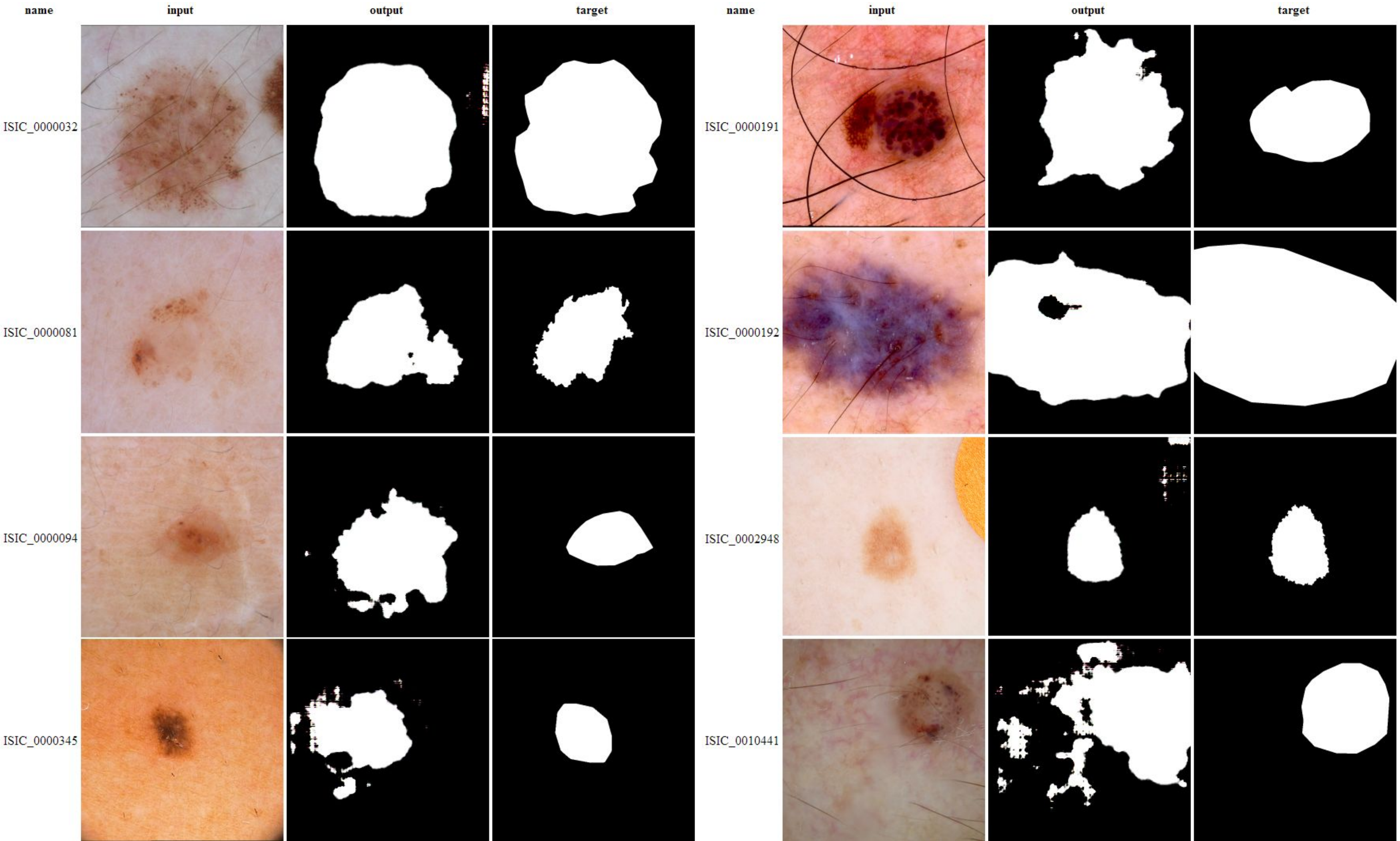
4. Lesion segmentation

- Competition **Skin Lesion Analysis Towards Melanoma Detection**
- Segment image - find lesion in the picture
- https://challenge.kitware.com/#challenge/560d7856cad3a57cfde481ba

- Solution: Analogue of pix2pix architecture




name
input
output
target
name
input
output
target
ISIC_0000006
ISIC_0000007
ISIC_0000021
ISIC_0000283
ISIC_0000034
ISIC_0000035
ISIC_0000038
ISIC_0000061



name
input
output
target
name
input
output
target
ISIC_0000032
ISIC_0000191
ISIC_0000081
ISIC_0000192
ISIC_0000094
ISIC_0002948
ISIC_0000345
ISIC_0010441

Conclusions

- For now GANs in healthcare were used for:
    - Generating new data
        - De-identification
        - Generating more data for further learning or other specific tasks
    - Anomaly detection in data
    - Feature extraction from images
    - Mapping new images to the latent space
    - Image segmentation
- GANs is an architecture that is rapidly developing

References

- Good tutorials from Ian Goodfellow:
  - https://arxiv.org/pdf/1701.00160.pdf
  - https://arxiv.org/pdf/1406.2661.pdf
  - https://youtu.be/RvgYvHyT15E
- Training hacks:
  - https://github.com/soumith/ganhacks
- GAN collections:
  - https://github.com/wiseodd/generative-models
  - https://github.com/hindupuravinash/the-gan-zoo

GlobalLogic®

Thank you!

Vladyslav Kolbasin
Lead Software Developer, Consultant
vladyslav.kolbasin@globallogic.com
ва