Medical data augmentation using GAN

Oleksandr Fedoruk, Aleksander Ogonowski, Konrad Klimaszewski

National Centre For Nuclear Research

WMLQ2022

September 14, 2022

European Funds Smart Growth
Republic of Poland
European Union European Regional Development Fund

Outline

Introduciton

EuroHPC PL

National Supercomputing Infrastructure for EuroHPC

The European High Performance Computing Joint Undertaking (EuroHPC JU) – joint initiative between the EU, European countries and private partner to develop a World Class Supercomputing Ecosystem in Europe

EuroHPC Joint Undertaking

https://eurohpc-ju.europa.eu/index_en

- Supercomputers
- GPU (Graphical Processor Units) computing
- Quantum computing
- Neuromorphic computing

EuroHPC PL – development of the specialized infrastructure for the exascale computations addressing the key challenges for the Polish society, scientific community and the economy.

EuroHPC PL

https://www.eurohpc.pl/

Consortium of 7 Polish institutions:
ACK Cyfronet AGH, PCSS, CI TASK, WCSS, NCBJ, IITiS PAN and CFT PAN.

CIS Świerk Computing Centre

NATIONAL CENTRE FOR NUCLEAR RESEARCH ŚWIERK

Software platform for quantum simulations and medical imaging

https://www.ncbj.gov.pl/en/aktualne/eurohpc-pl-national-supercomputing-infrastructure-eurohpc

European Funds Smart Growth

Republic of Poland

European Union European Regional Development Fund

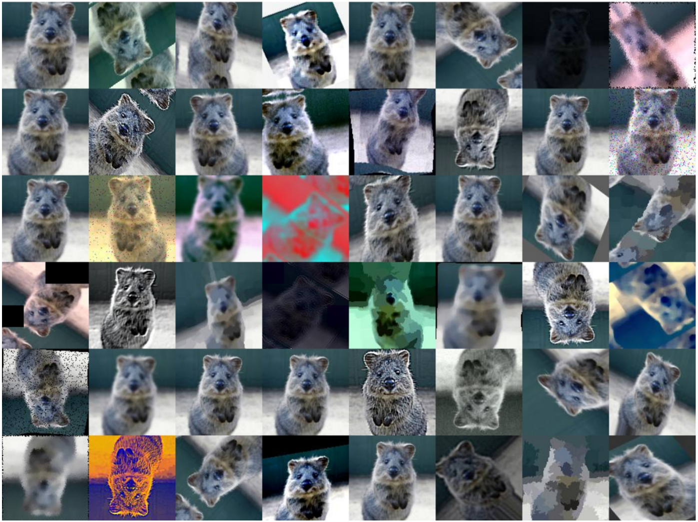
Goal of the research

The goal of the research is to verify if Generative Adversarial Networks-based data augmentation is a suitable method to achieve better performance of medical image data classification in comparison with classical image data augmentation.

Problems with medical data:

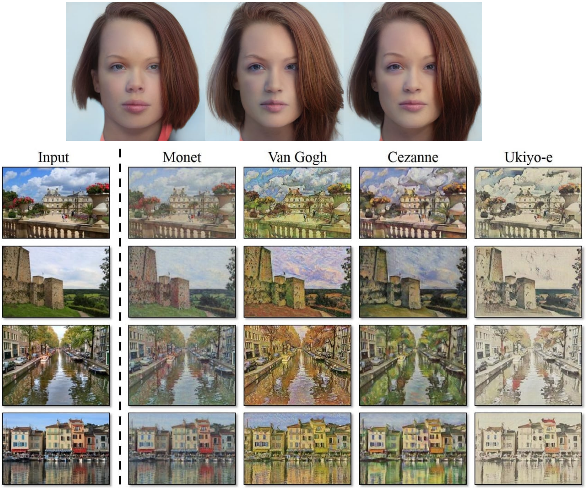
- Medical data is usually protected as a private information (Medical privacy, GDPR, etc.).
- Some medical procedures are too expensive to perform and gather a lot of data
- Some diseases are too rare to gather a lot of data

Current solution - classic augmentation

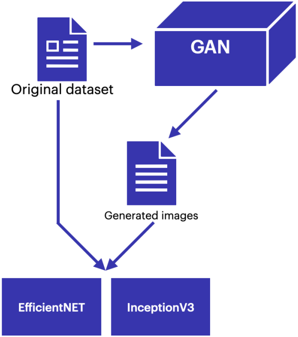


Example of image augmentation

GAN augmentation



Some example of GANs usage



Dataset

Dataset

Research is based on the "**COVID-19 Radiography Database**"[1] available at Kaggle.com.

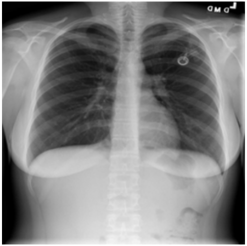
The dataset incorporates x-ray scans from multiple public sources and contains:

- 3616 images of COVID-19.
- 6012 images of Lung Opacity.
- 1345 images of Viral Pneumonia.
- 10192 images of healthy lungs.

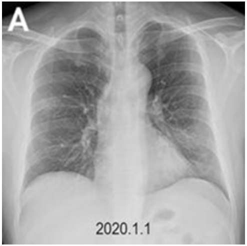
For all images lung masks are available.

[1] www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

Sample images from dataset



Healty/Normal



COVID-19

Lung Opacity

Viral pneumonia

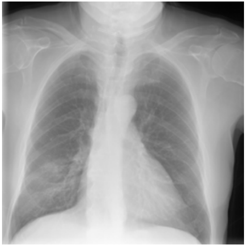
Lungs Mask

Data preprocessing

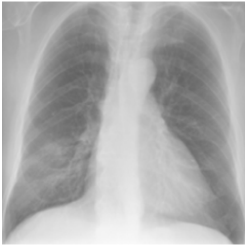
Data preprocessing

Data preprocessing consists of the three steps:

1. cropping the images to rectangular that contains lung only,
2. resizing new images to 256x256 pixels,
3. normalising images to [0; 1] scale



Original COVID-19 image

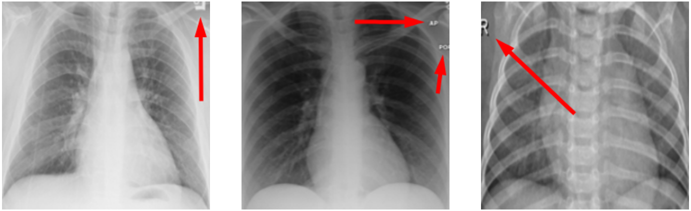


Same image cropped and resized

Data preprocessing

Many images in the dataset contain some kind of annotations or marks made by medical personnel or the scanner itself.

After cutting and resizing images we manually removed all images containing such elements.

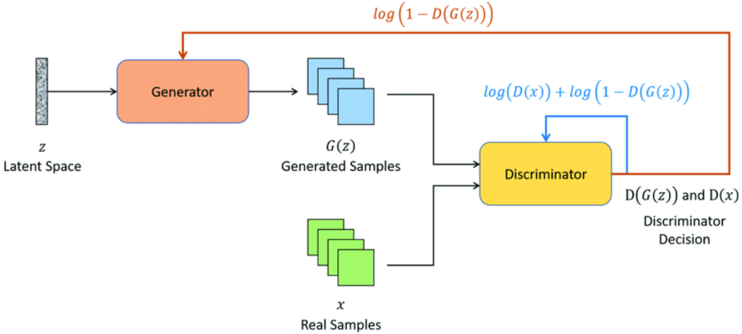


Examples of annotations and marks on cut and resized images

Generative Adversarial Networks

Generative Adversarial Networks

Generative Adversarial Network (GAN) - is a class of deep learning framework designed by Ian Goodfellow and his colleagues. [2]



Typical Generative Adversarial Networks architecture [3]

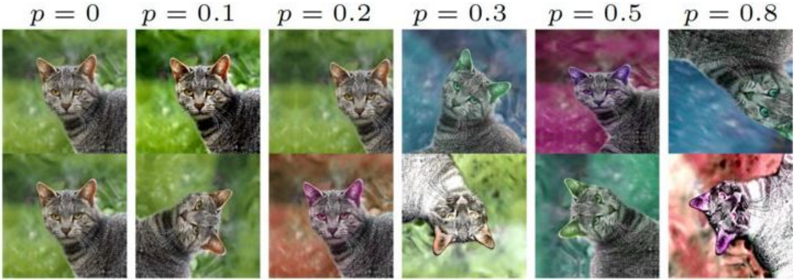
---

[3] arXiv:1406.2661 [stat.ML]

[3] https://www.researchgate.net/figure/Typical-Generative-Adversarial-Networks-GAN-architecture$_fig2_3$49182009[$accessed$11Sep, 2022]

StyleGAN-2 + ADA

In December 2019 an NVIDIA team has introduced better version of GAN architecture - StyleGAN[4].
Later, in June 2020 they introduced adaptive discriminator augmentation mechanism that significantly stabilises training with limited data. [5]



[5] arXiv:1912.04958 [cs.CV]
[5] arXiv:2006.06676 [cs.CV]

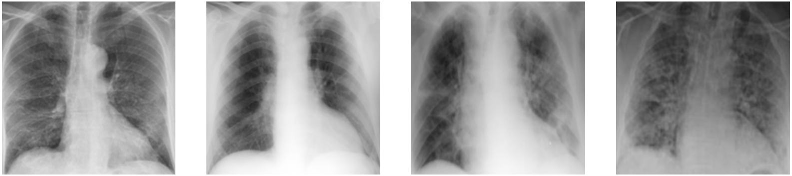
GAN training with COVID data

GAN training process

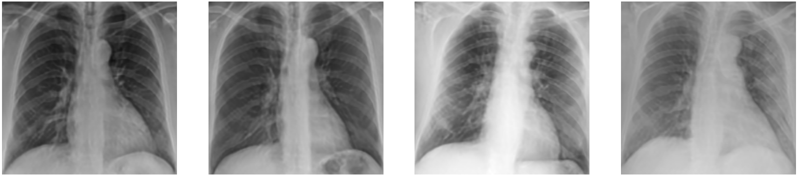
In order to have better control when comparing different scenarios, we have taken 200 COVID and 200 healthy scans to store them separately as so-called "test dataset". The test dataset wasn't used while training GAN and classification networks so it represented new never seen before images. Final dataset was containing:

- 2949 original COVID images in training set
- 9992 original healthy images in training set
- 2000 GAN generated COVID images in training set
- 200 original COVID images in test set
- 200 original healthy images in test set

GAN-generated vs Original images similarity



Examples of real COVID X-Ray images



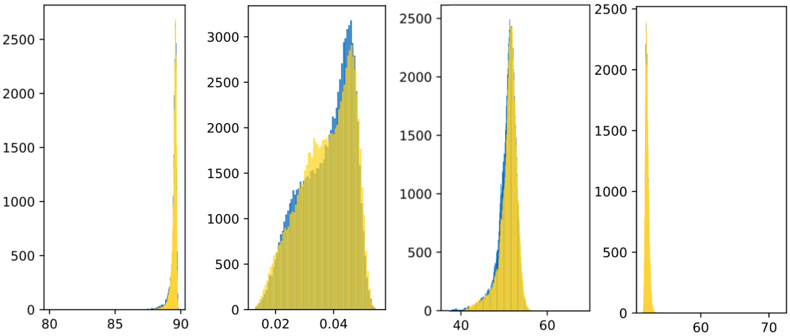
Examples of GAN-generated COVID X-Ray images

Image comparison metrics

In order to verify that generated image are similar to it's original sources, we've applied several image comparison metrics to generated and originals sets.

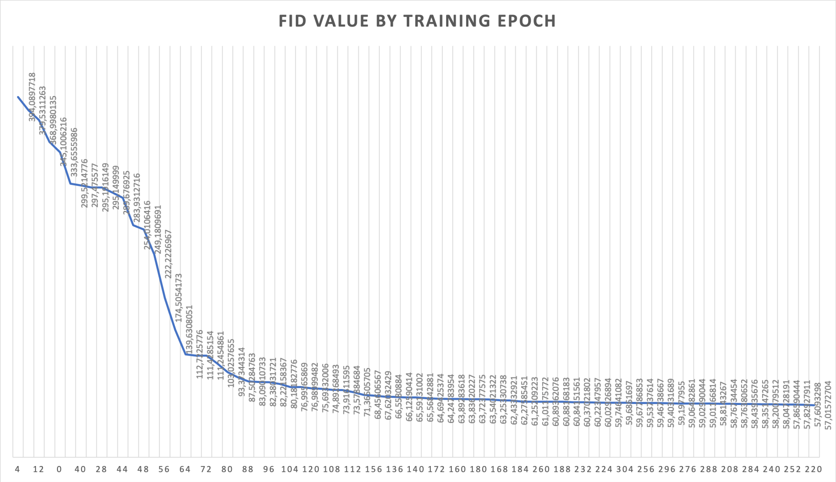
- RMSE - root mean square error
- SRE - signal to reconstruction error ratio
- SAM - spectral angle mapper
- PSNR - peak signal-to-noise ratio
- FID - fréchet inception distance

Image comparison metrics



SAM RMSE SRE PSNR

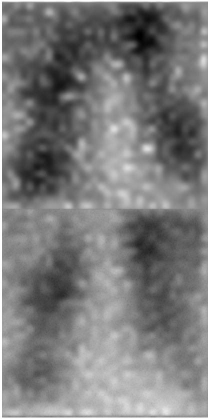
FID



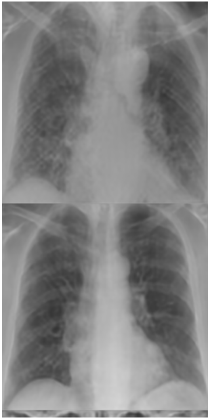
FID values decreasing with each epoch of Style-GAN training

FID - examples

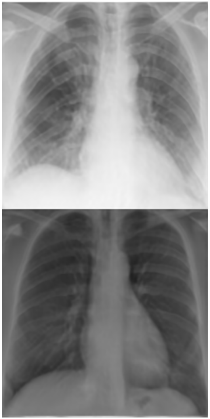
Comparing generated images to original dataset



FID $\approx$ 370



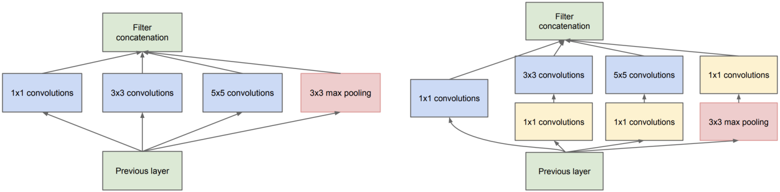
FID $\approx$ 73



FID $\approx$ 56

Classification Networks

Inception V3

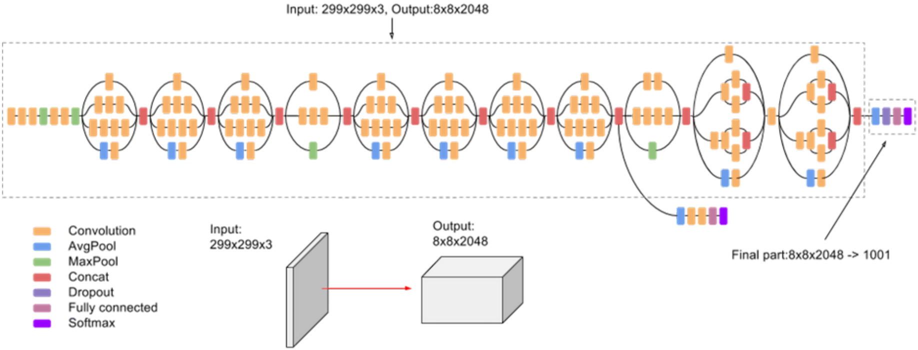


(a) Inception module, naïve version (b) Inception module with dimension reductions

Figure 2: Inception module

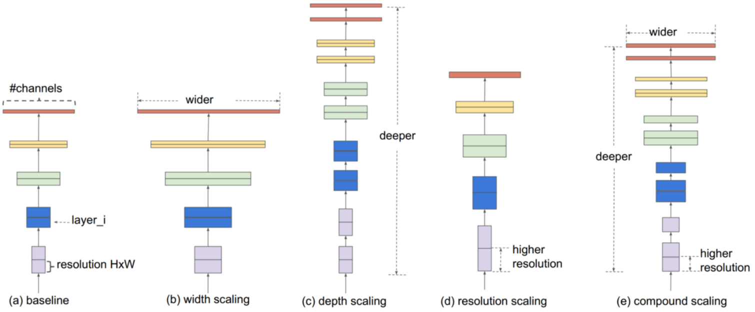
Architecture of inception block

Inception V3



Architecture of InceptionV3

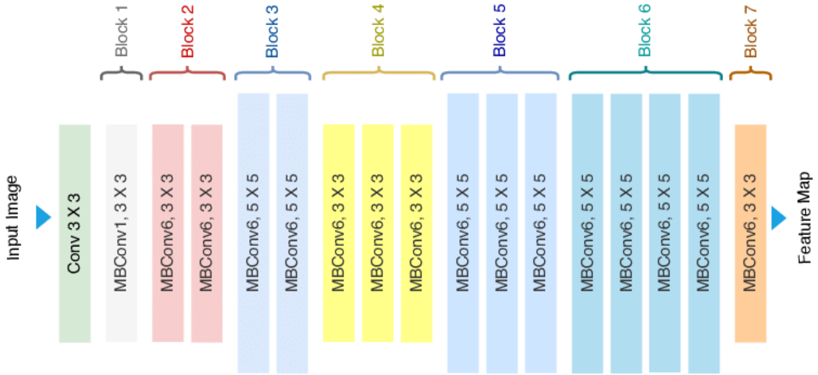
EfficientNet



Different scaling methods vs. Compound scaling [6]

[6] arXiv:1905.11946 [cs.LG]

EfficientNet



Architecture of EfficientNetB0 (baseline model)

Training process

We have trained 2 networks architectures with 3 different augmentation approaches.

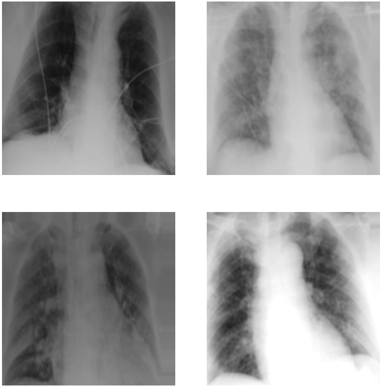
- No augmentation at all
- Classical augmentation
- GAN augmentation

All networks were trained with 6 epochs (with batch size equal to 32) per training. Epoch with best validation accuracy was picked as a training result.
Overall quality of each augmentation approach is described as accuracy calculated on testing set (400 images, 200 images per class).

Classical augmentation

Parameters of classical augmentation

- Validation split=0.1
- Rotation range = 5
- Width shift range = 0.05
- Height shift range = 0.05
- Shear range = 0.05
- Zoom range = 0.1
- Horizontal flip = False
- Brightness range = [0.6,1.4]



Examples of augmented images.

Results, summary and outlook

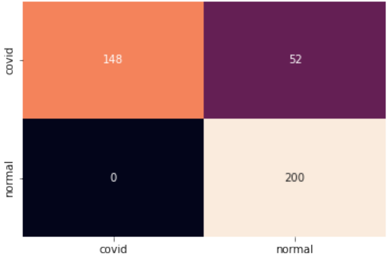
Results, no augmentation

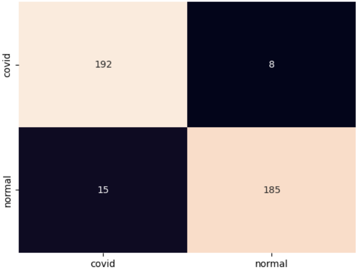
InceptionV3

- Accuracy - 0.875
- True Positive Rate - 0.74
- True Negative Rate - 1
- False Positive Rate - 0.26
- Mean accuracy from 4 training runs - 0.8875

EfficientNet

- Accuracy - 0.9425
- True Positive Rate - 0.96
- True Negative Rate - 0.925
- False Positive Rate - 0.04
- Mean accuracy from 4 training runs - 0.9475

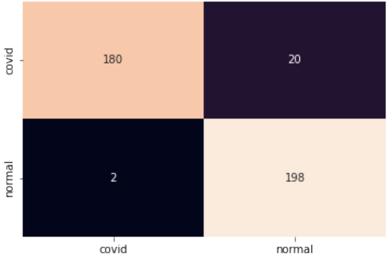




Results, classic augmentation

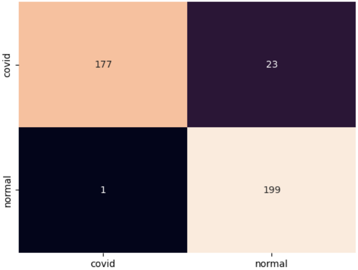
InceptionV3

- Accuracy - 0.945
- True Positive Rate - 0.9
- True Negative Rate - 0.99
- False Positive Rate - 0.1
- Mean accuracy from 4 training runs - 0.949375



EfficientNet

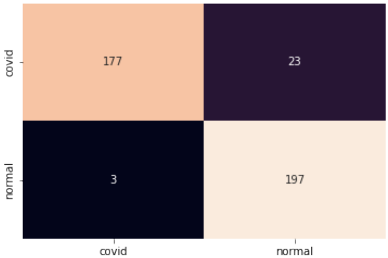
- Accuracy - 0.94
- True Positive Rate - 0.885
- True Negative Rate - 0.995
- False Positive Rate - 0.115
- Mean accuracy from 4 training runs - 0.904375



Results, GAN augmentation

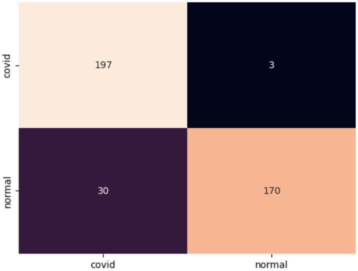
InceptionV3

- Accuracy - 0.935
- True Positive Rate - 0.885
- True Negative Rate - 0.985
- False Positive Rate - 0.115
- Mean accuracy from 4 training runs - 0.885



EfficientNet

- Accuracy - 0.9175
- True Positive Rate - 0.985
- True Negative Rate - 0.85
- False Positive Rate - 0.015
- Mean accuracy from 4 training runs - 0.905625



Results

Comparison of accuracy values

	InceptionV3	EfficientNet
No augmentation	0.8875	0.9475
Classic augmentation	0.949375	0.904375
GAN augmentation	0.885	0.905625

Summary

- We've tested InceptionV3 and EfficientNet with transfer learning to classify COVID-19 lungs scan
- We've verified the classical augmentation impact on classification quality
- We've demonstrated the results of GAN augmentation and it's impact on classification quality

Outlook

- Verify impact of different data augmentation pipelines for 4 class image classification (COVID, healthy, lungs opacity, viral pneumonia)
- Introduce a new data augmentation pipeline (classical and GAN combined)
- Train Style-GAN using it's style transfer feature so it would allow to populate every dataset with one GAN