# Anomaly Detection with GANs

## Application in medical image analysis

**Thomas Schlegl**

**Vienna Deep Learning Meetup**

October 29, 2019

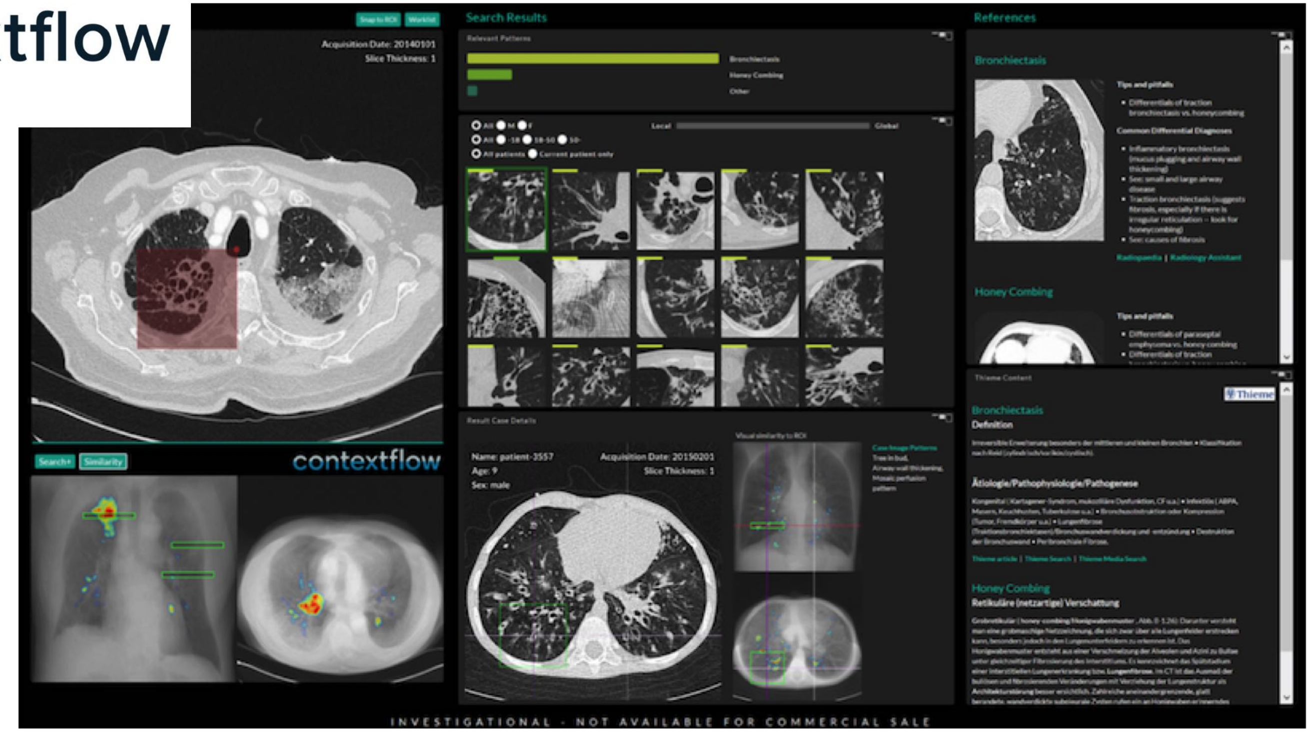
# About me - Thomas Schlegl

MEDICAL UNIVERSITY OF VIENNA

CIR

OPTIMA
Ophthalmic Image Analysis

contextflow

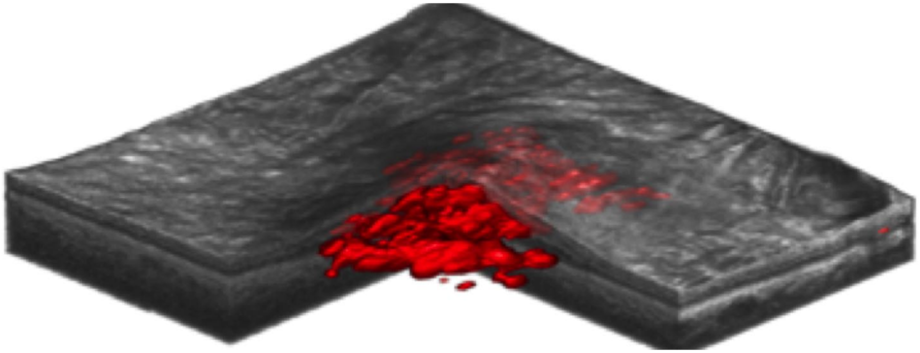


# Overview

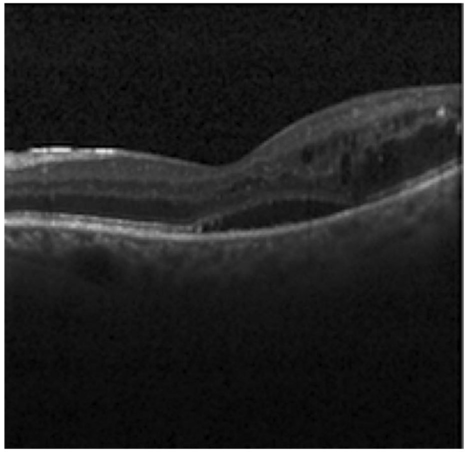
1. Clinical background
2. Motivation: Anomaly detection
3. Generative Adversarial Network (GAN)
4. (f-)AnoGAN: Anomaly Detection with GANs
5. Conclusion

# Clinical background

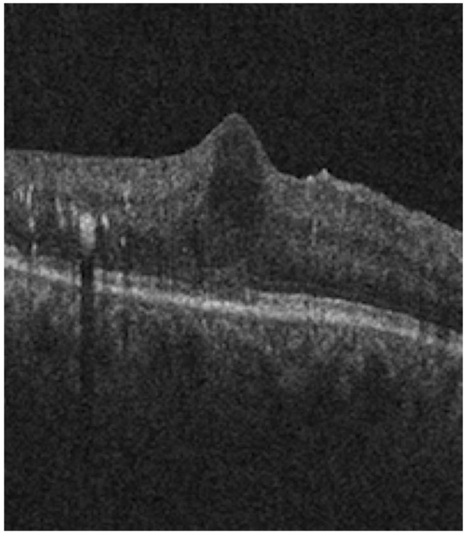
# 3-D Retinal imaging



- Spectral-domain optical coherence tomography (SD-OCT)
- High resolution 3D imaging technique



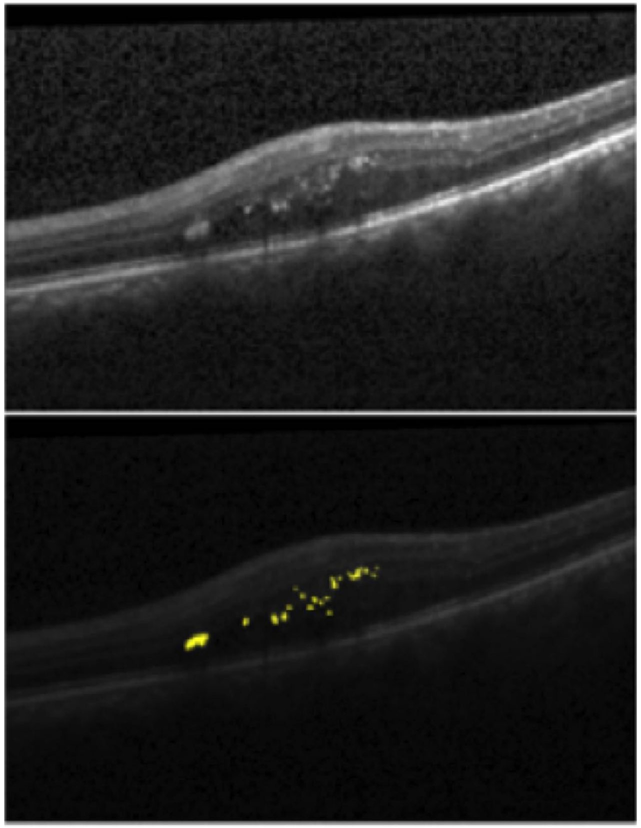
Spectralis scanner
$(496 \times 512) \times 49$



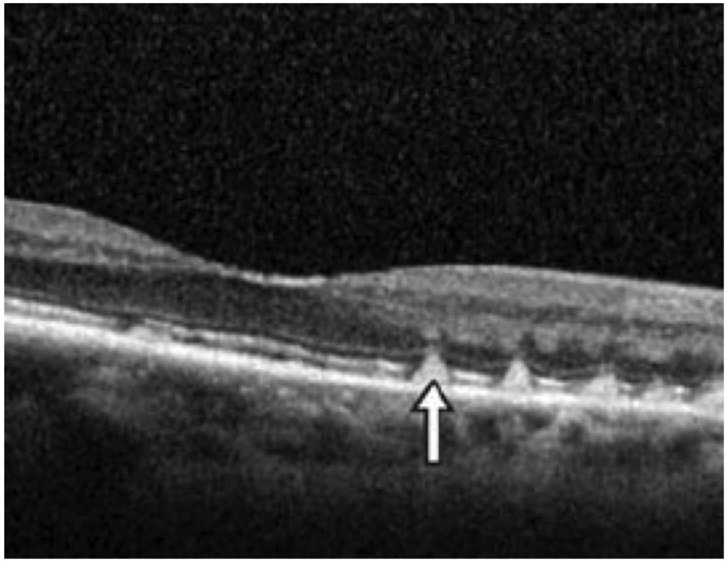
Cirrus scanner
$(1024 \times 512) \times 128$

# Imaging patterns in retinal OCT

## Hyperreflective foci (HRF)



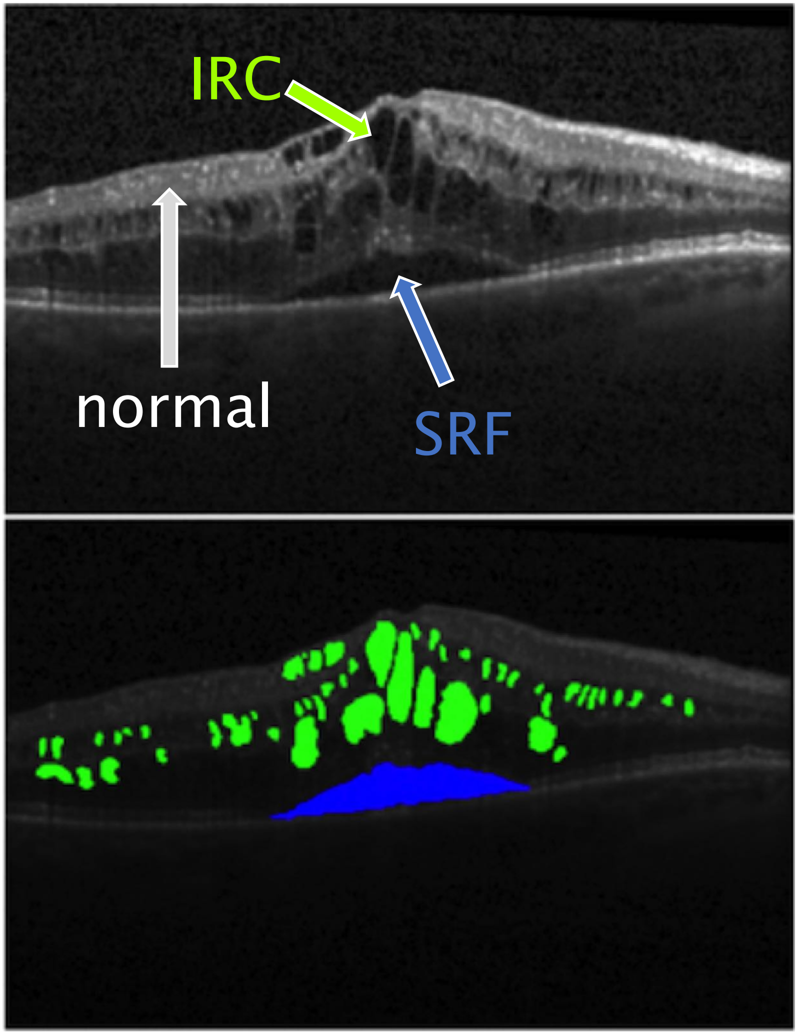
## Pseudodrusen



Yonekawa & Kim, 2015

# Imaging patterns in retinal OCT

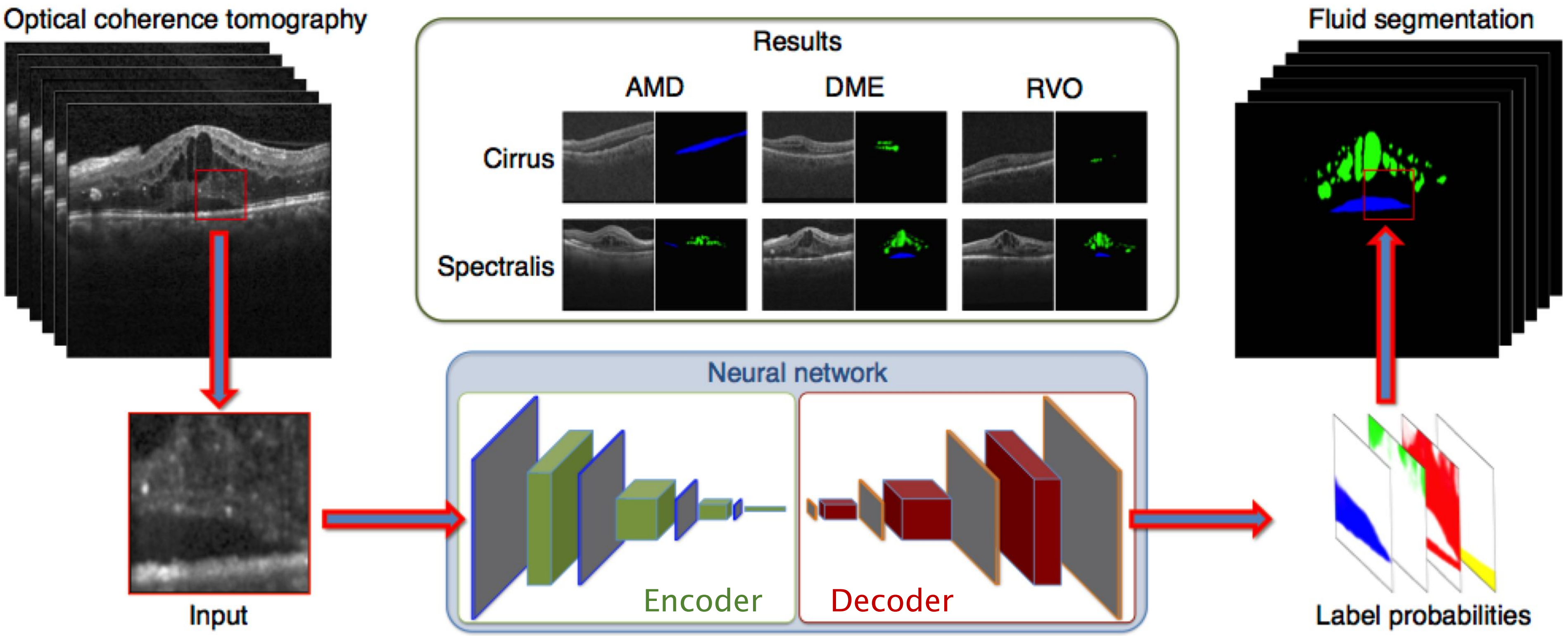
- Macular fluid:
  - Intraretinal Cystoid Fluid (**IRC**)
  - Subretinal Fluid (**SRF**)
- Normal (non-fluid) retinal tissue
- Relevant diseases:
  - Age-related Macular Degeneration (AMD)
  - Diabetic Macular Edema (DME)
  - Retinal Vein Occlusion (RVO)



# Motivation

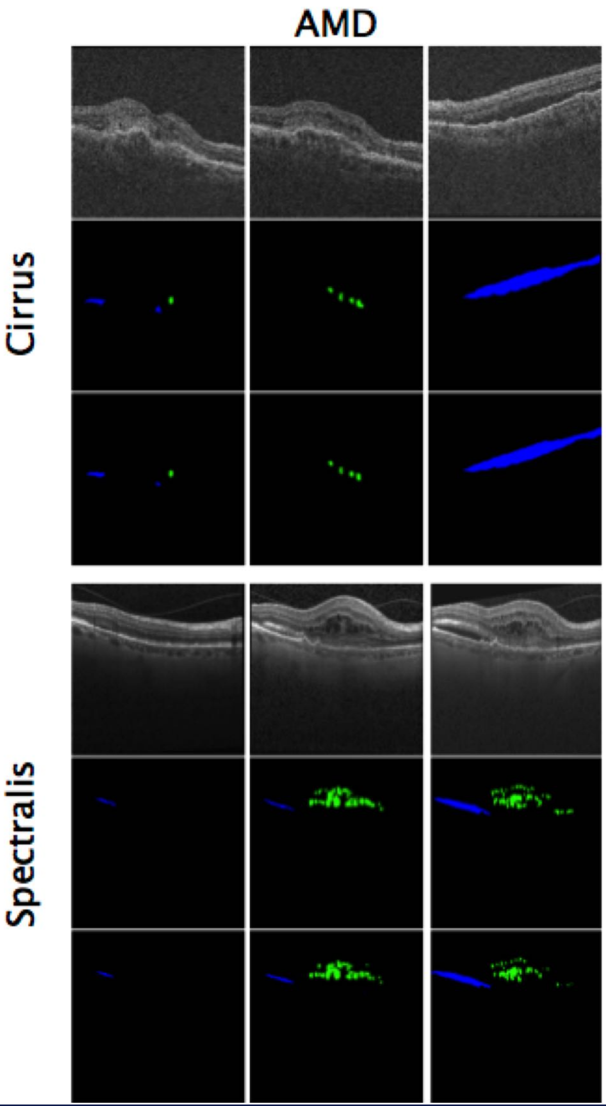
## Anomaly detection

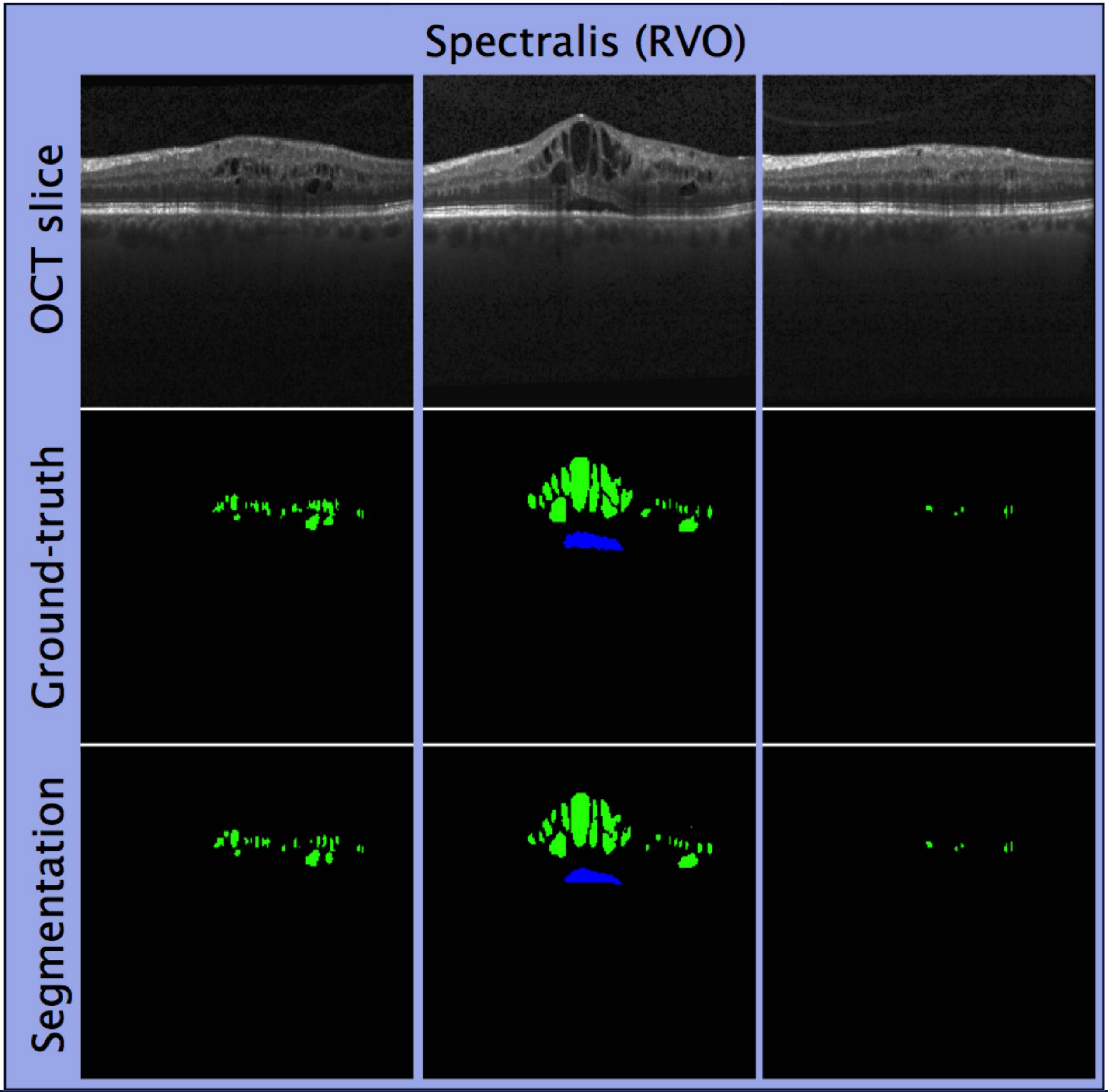
# Supervised learning




Schlegl et al. (2018). Fully automated detection and quantification of macular fluid in OCT using deep learning. Ophthalmology, 125(4), 549-558.

# Segmentation

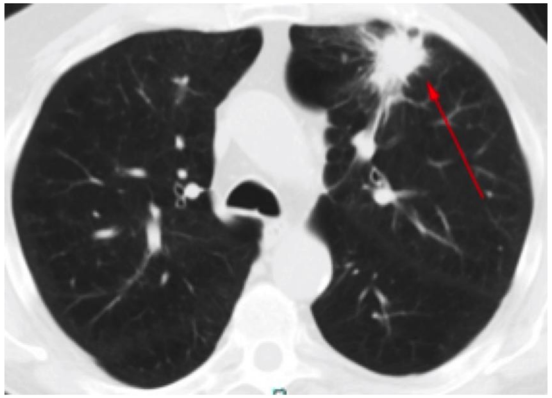


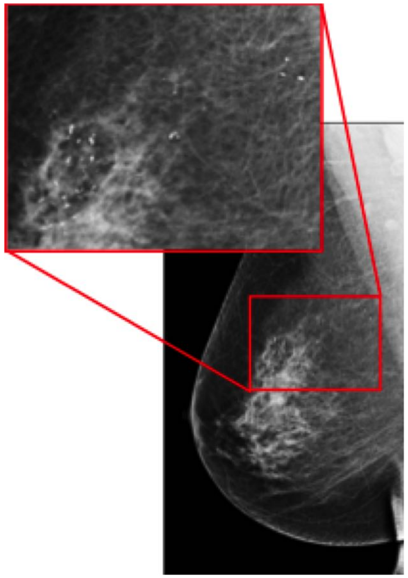



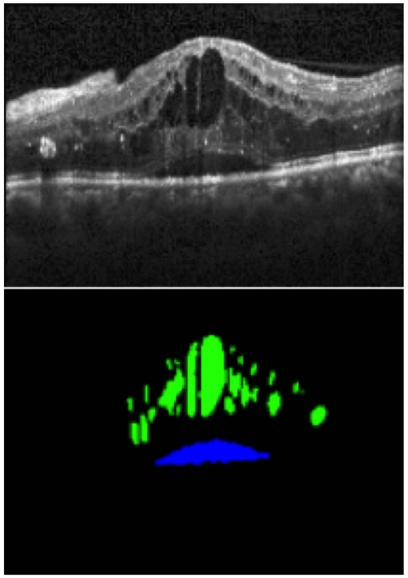
# Imaging biomarkers

- Diagnosis
- Disease management
- Supervised learning:
  - Accurate detection

- Limitations of supervised approaches:
  - Requires annotations
  - Restriction to a vocabulary of known markers







# Anomaly detection

- Anomaly:
  - Deviating (not “normal”) sample
  - Not drawn from the distribution of normal samples

- Example domains:
  - Credit-card fraud detection in **finance**
  - Radio frequency anomaly detection in **wireless networks**
  - Lesion detection and novel biomarker identification in **medical image analysis**

# Anomaly detection approaches

- **Statistical approaches** (Nguyen & Goulet, 2017)
- **Density-based** anomaly detection (Zhang et al., 2018)
- **Clustering-based** anomaly detection (Alguliyev et al., 2017)
- **Graph-based** anomaly detection ( Akoglu et al., 2015)
- **Support vector machine** (SVM)-based anomaly detection (Erfani et al., 2016; Seeböck et al., 2018)

# (f-)AnoGAN: Anomaly detection with GANs

[Initial work: *AnoGAN*]

Unsupervised anomaly detection with generative adversarial networks to guide marker discovery.

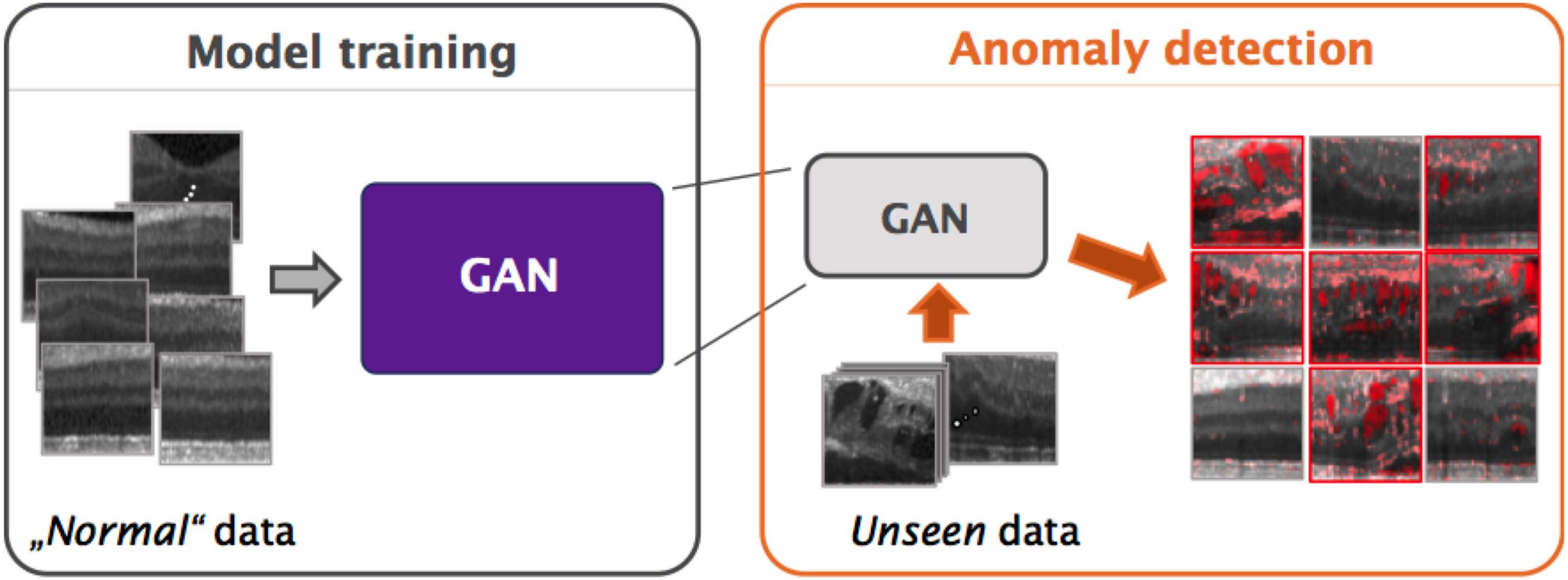
Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., Langs, G. (2017). In International Conference on Information Processing in Medical Imaging (pp. 146-157). Springer, Cham. arXiv preprint arXiv:1703.05921.

---

f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks.

Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., Schmidt-Erfurth, U. (2019). Medical Image Analysis. DOI: https://doi.org/10.1016/j.media.2019.01.010

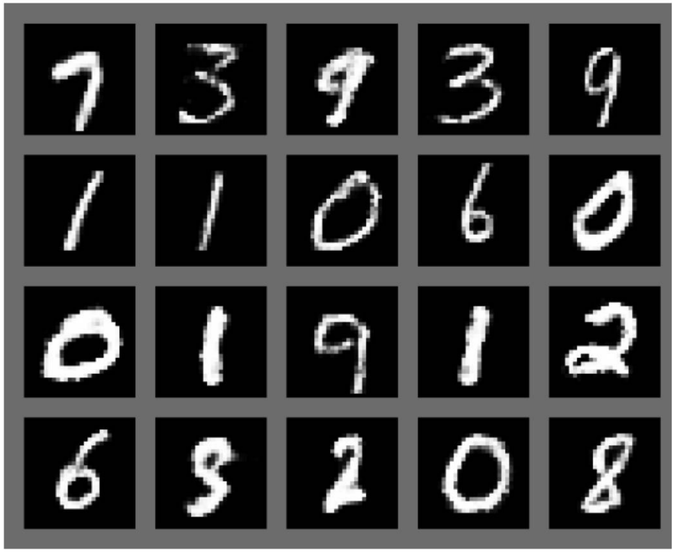
# Anomaly detection framework



# Generative Adversarial Network

## GAN

# Generative adversarial network (GAN)



Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

# Generative adversarial network (GAN)

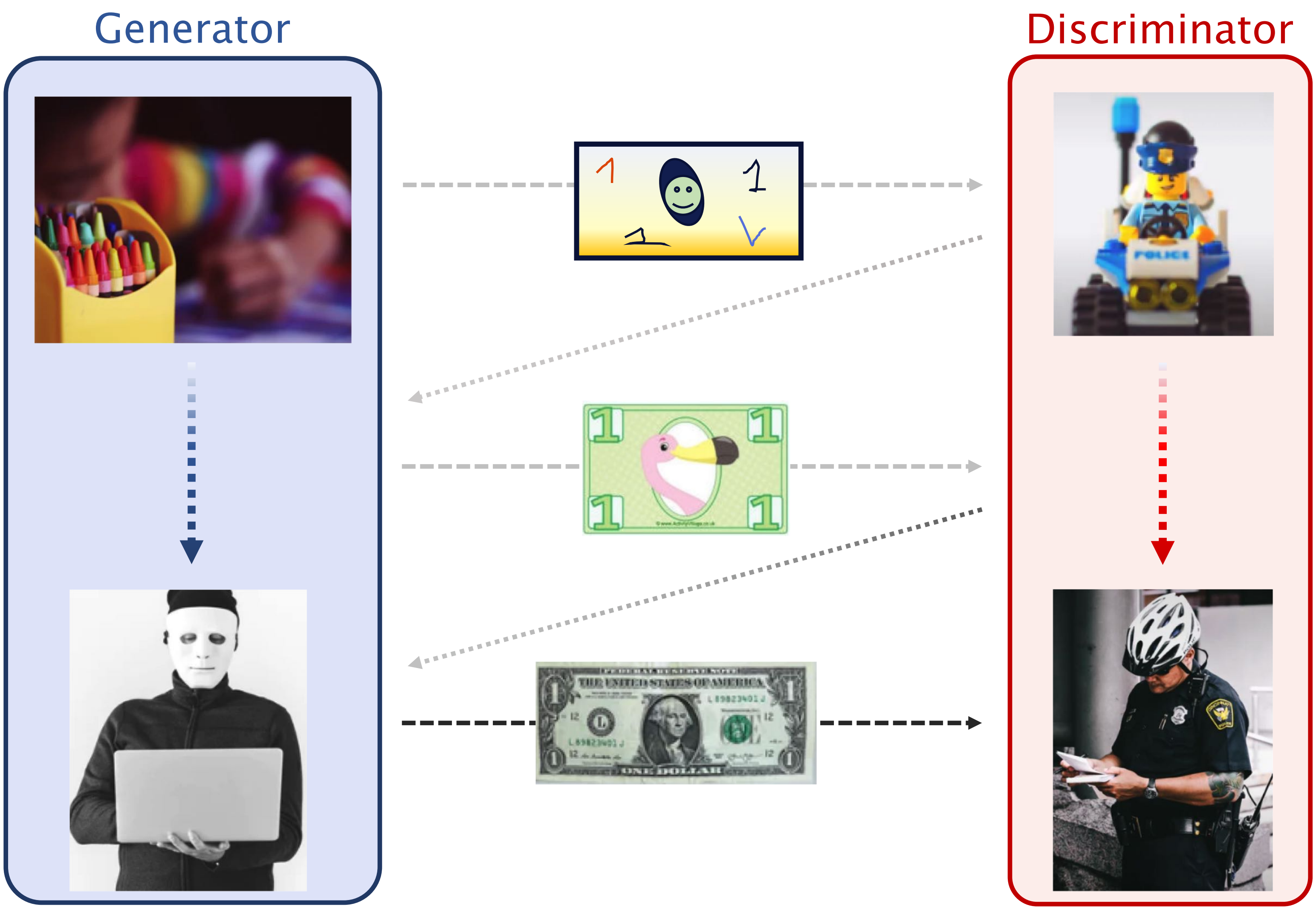
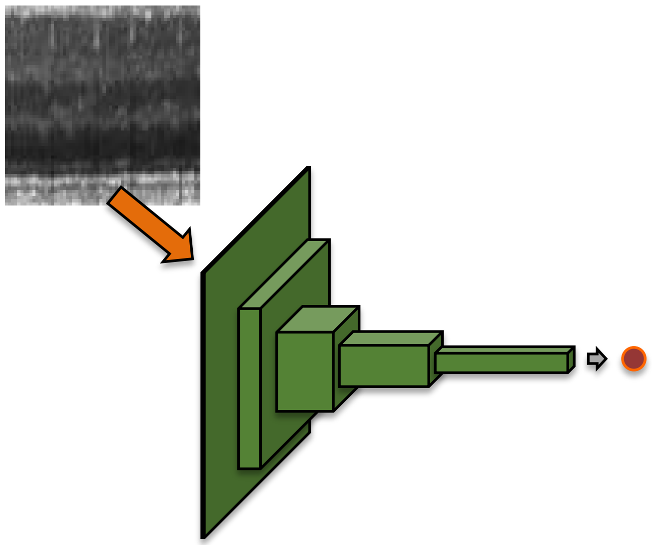



Image sources: https://unsplash.com/search/photos/photos
https://www.activityvillage.co.uk/fun-money-banknotes-1

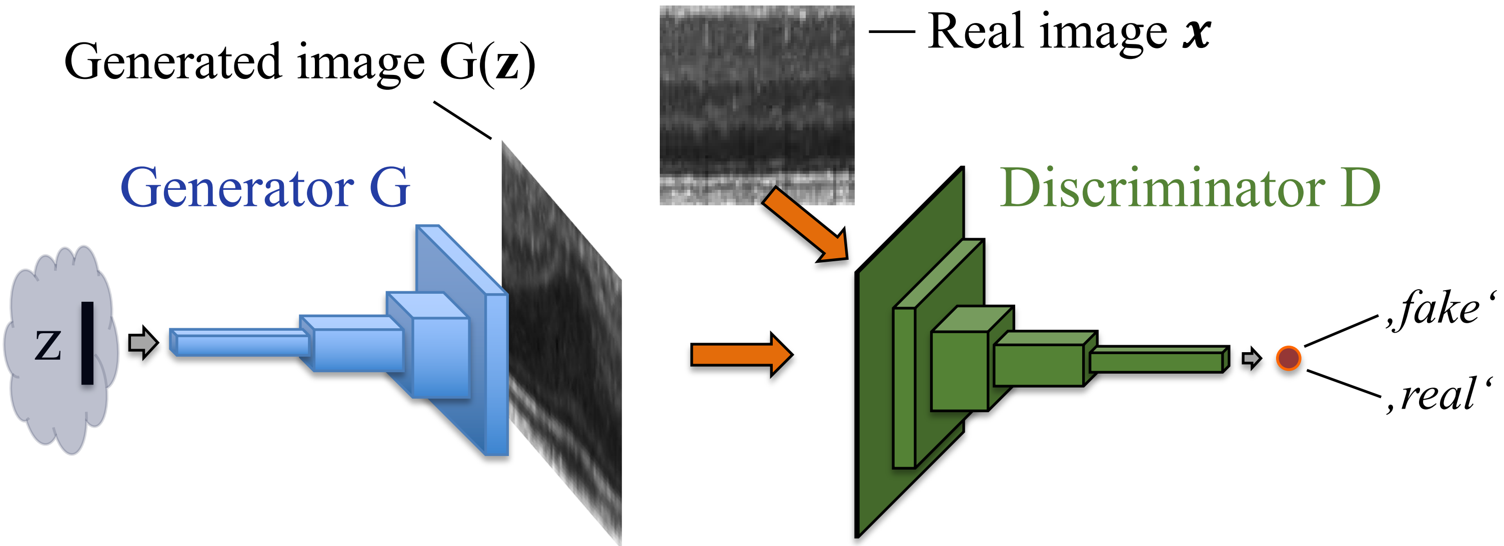
# Binary classifier training



- Supervised learning: needs *ground-truth* targets $y$

# Generative adversarial network (GAN) training

[Goodfellow et al., 2014]



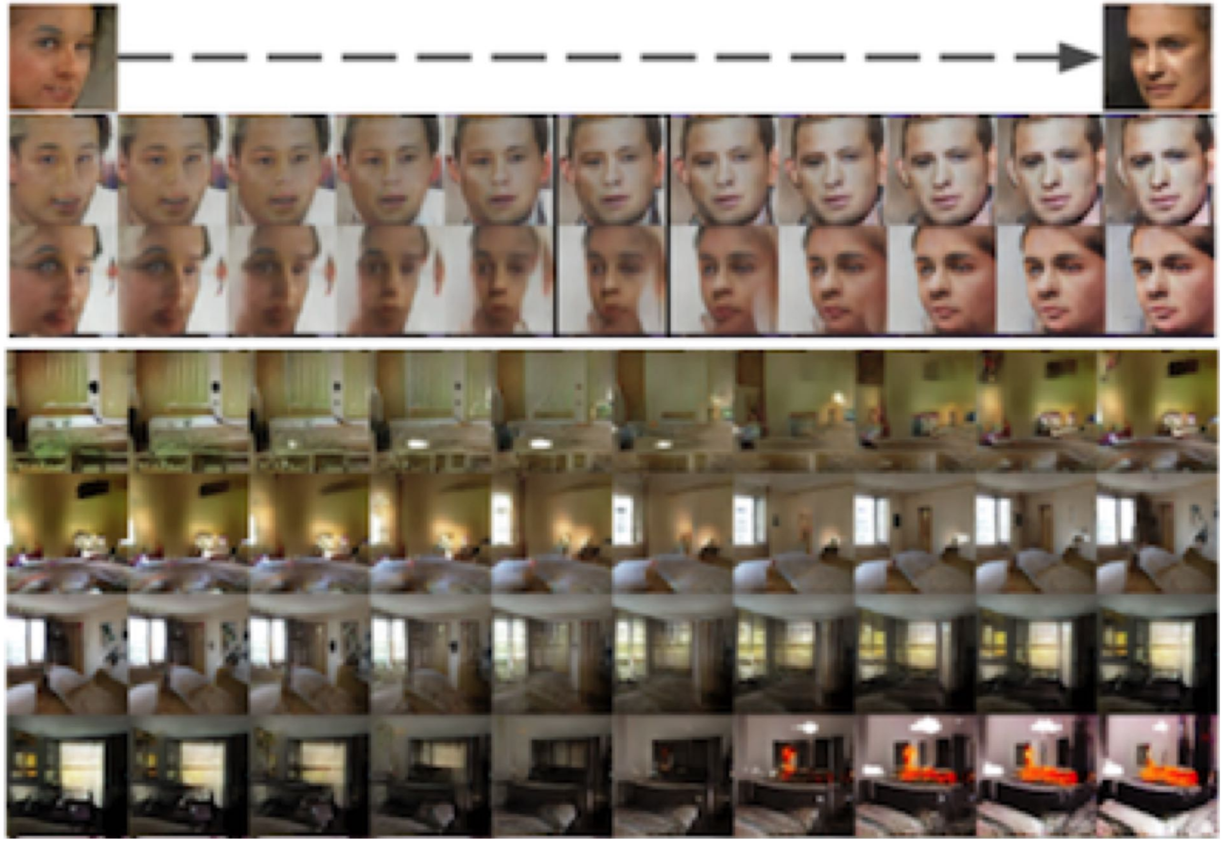

- GAN training:

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

# DCGAN

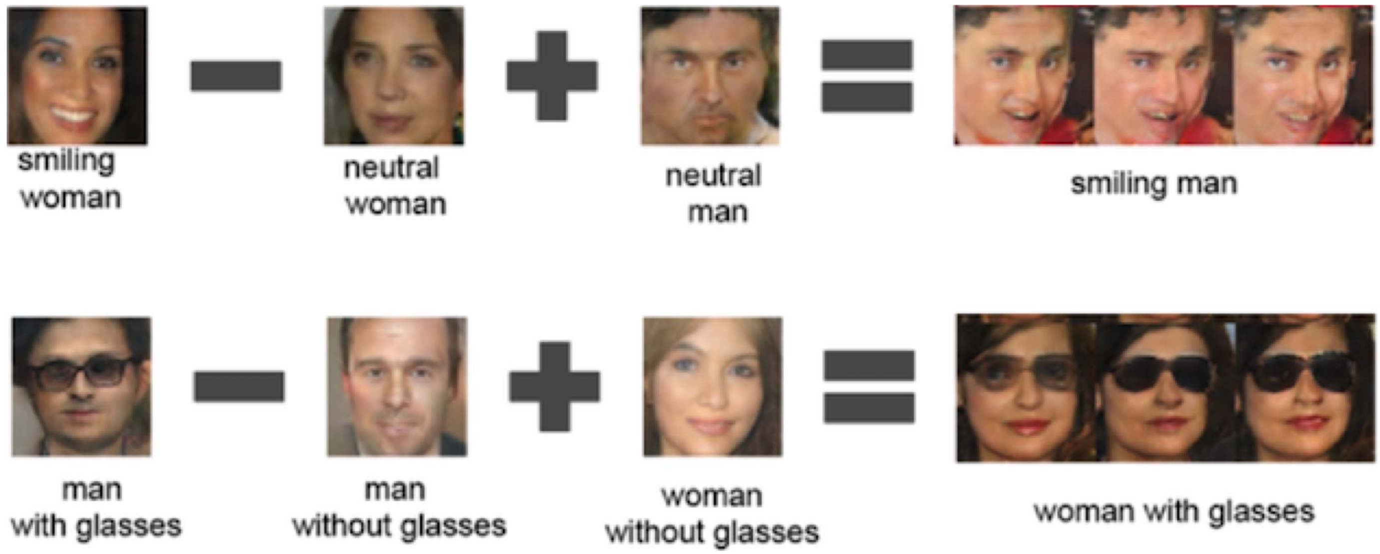
Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. ICLR 2016. arXiv preprint arXiv:1511.06434.

# DCGAN:
# Interpolation in latent space



Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

# DCGAN:
# Vector arithmetic for visual concepts




Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

# Plug & Play Generative Networks


Generated high-resolution (227x227px) images

Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A., & Yosinski, J. (2017). Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space. In CVPR 2017 (Vol. 2, No. 5, p. 7).

# StackGAN

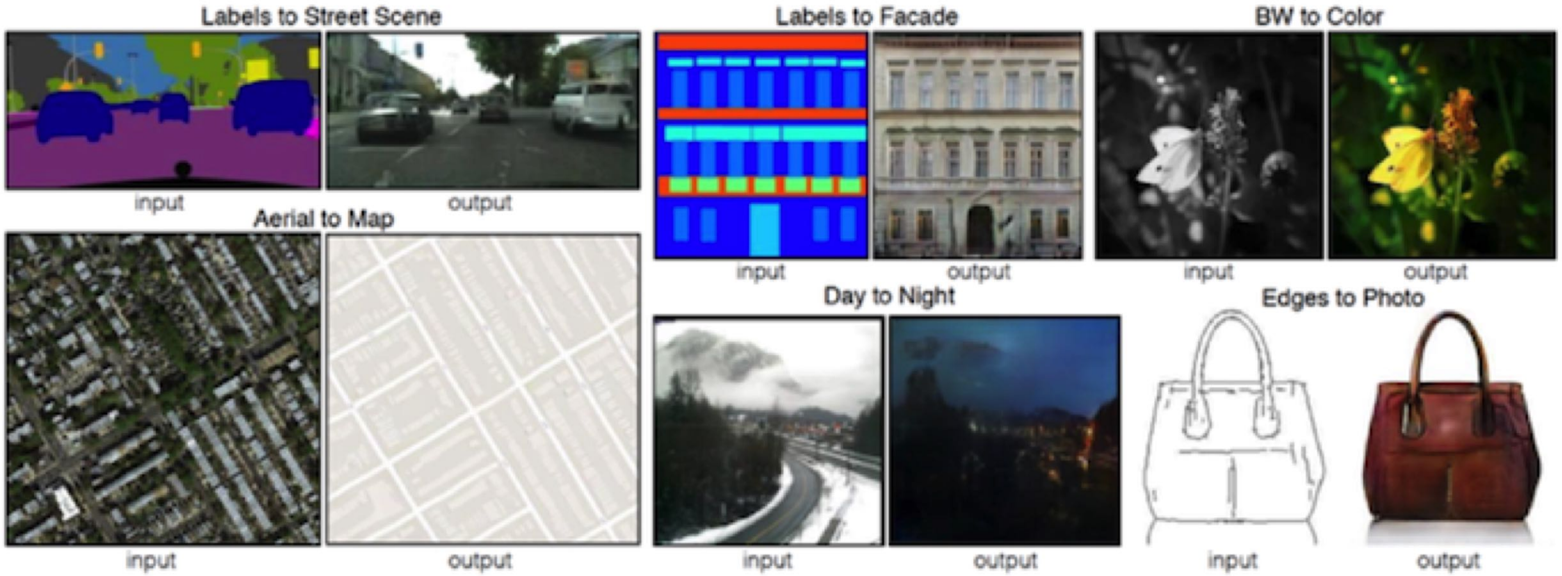
This bird is white with some black on its head and wings, and has a long orange beak

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

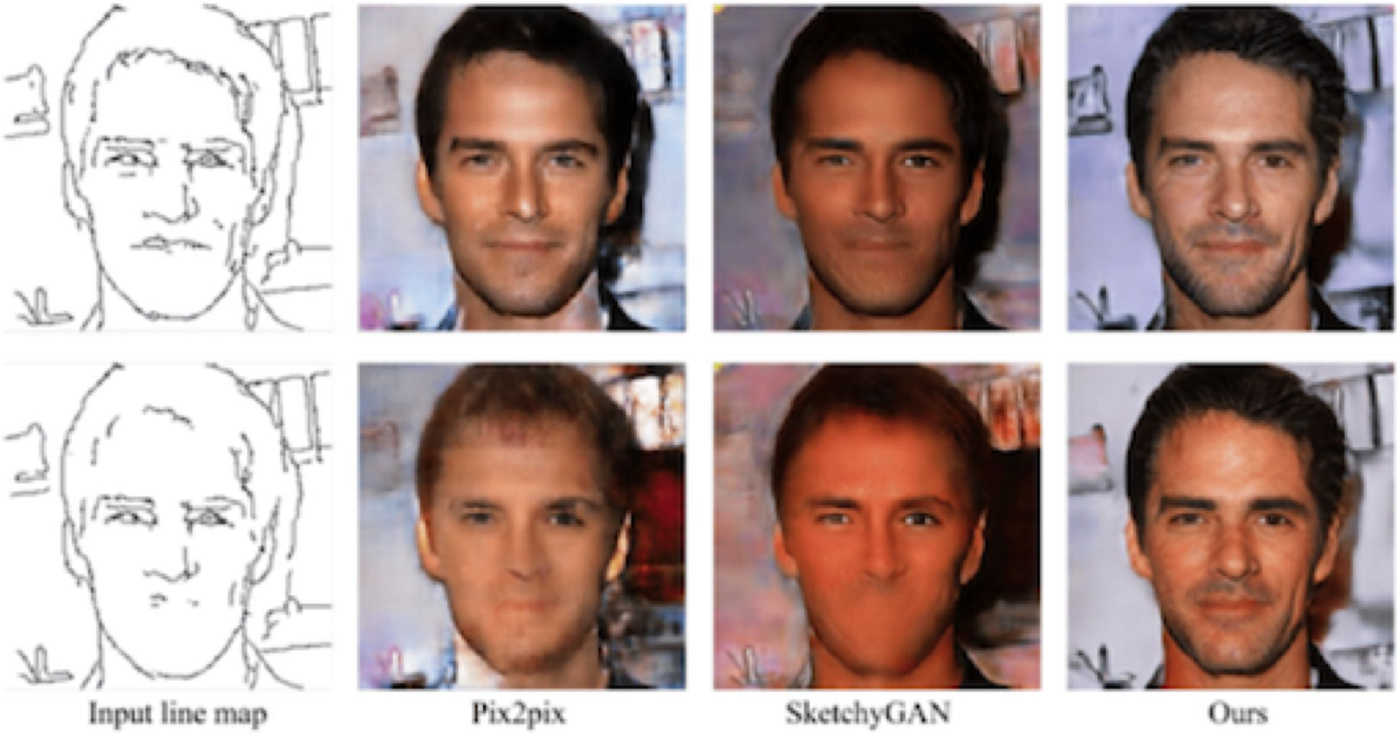
Zhang, H., Xu, T., Li, H., Zhang, S., Huang, X., Wang, X., & Metaxas, D. (2016). StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. arXiv preprint arXiv:1612.03242.

# “pix2pix” GAN



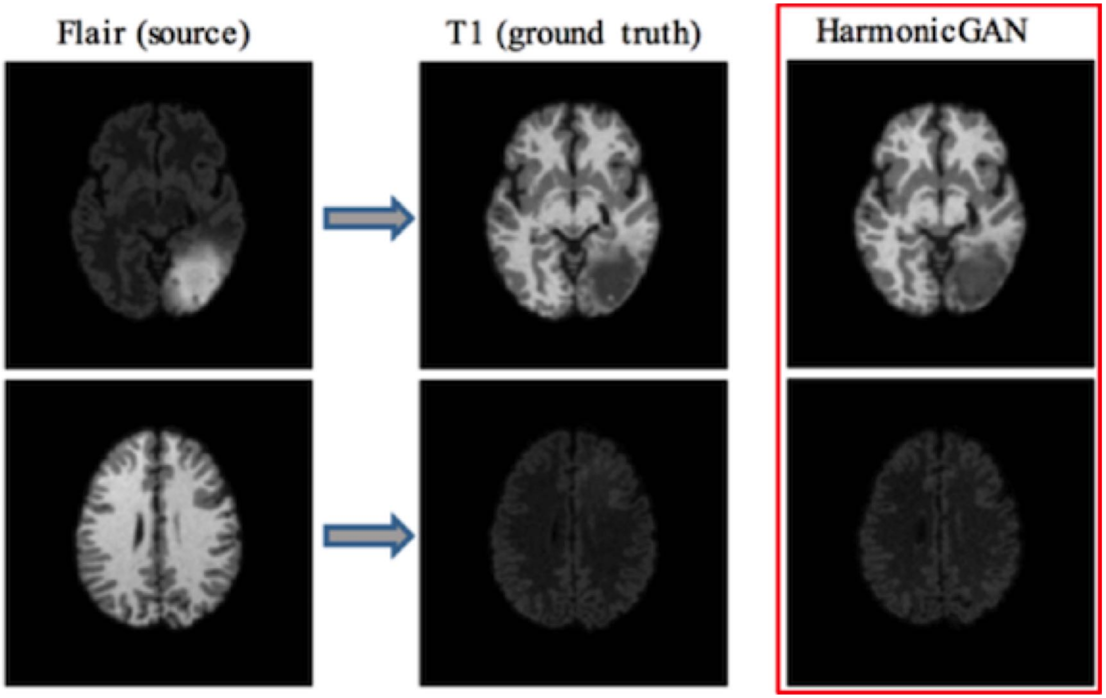

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. CVPR 2017. arXiv preprint.

# Conditional Self-Attention Generative Adversarial Network (CSAGAN)




Li, Y., Chen, X., Wu, F., & Zha, Z. J. (2019). LinesToFacePhoto: Face Photo Generation from Lines with Conditional Self-Attention Generative Adversarial Network. arXiv preprint arXiv:1910.08914.

# HarmonicGAN




Zhang, R., Pfister, T., Li, J. (2019). Harmonic unpaired image-to-image translation. ICLR 2019.

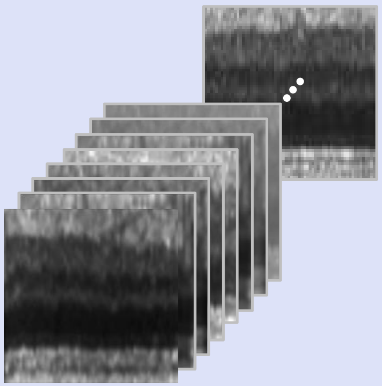
# AnoGAN

## Anomaly detection with GANs

# Anomaly detection with AnoGAN

**1 Training the model**

Deep Convolutional Generative Adversarial Network (DCGAN)



**Model**

Generator G

Discriminator D

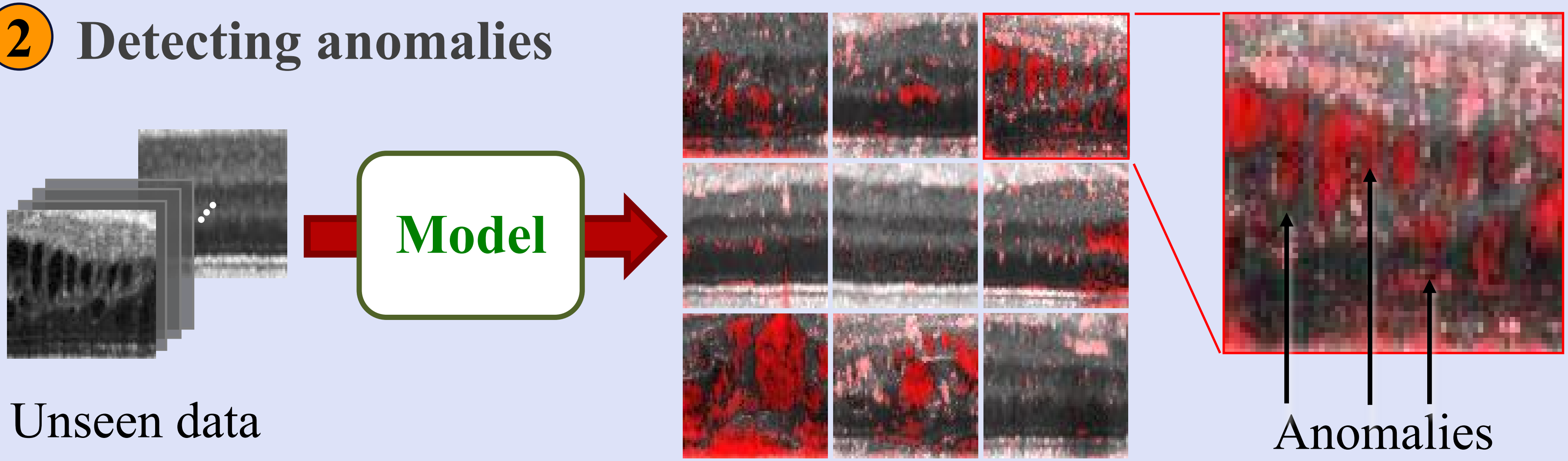
Normal (‚*healthy*') data

**2 Detecting anomalies**

# Anomaly detection with AnoGAN

**1 Training the model (DCGAN)**

**2 Detecting anomalies**



# 2 Detecting anomalies

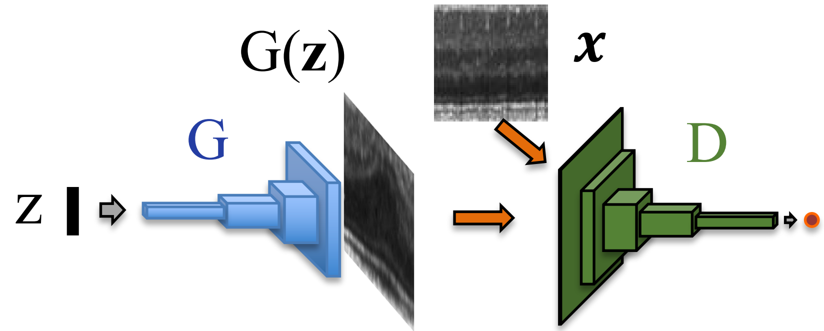
1. ***Normality* mapping:**

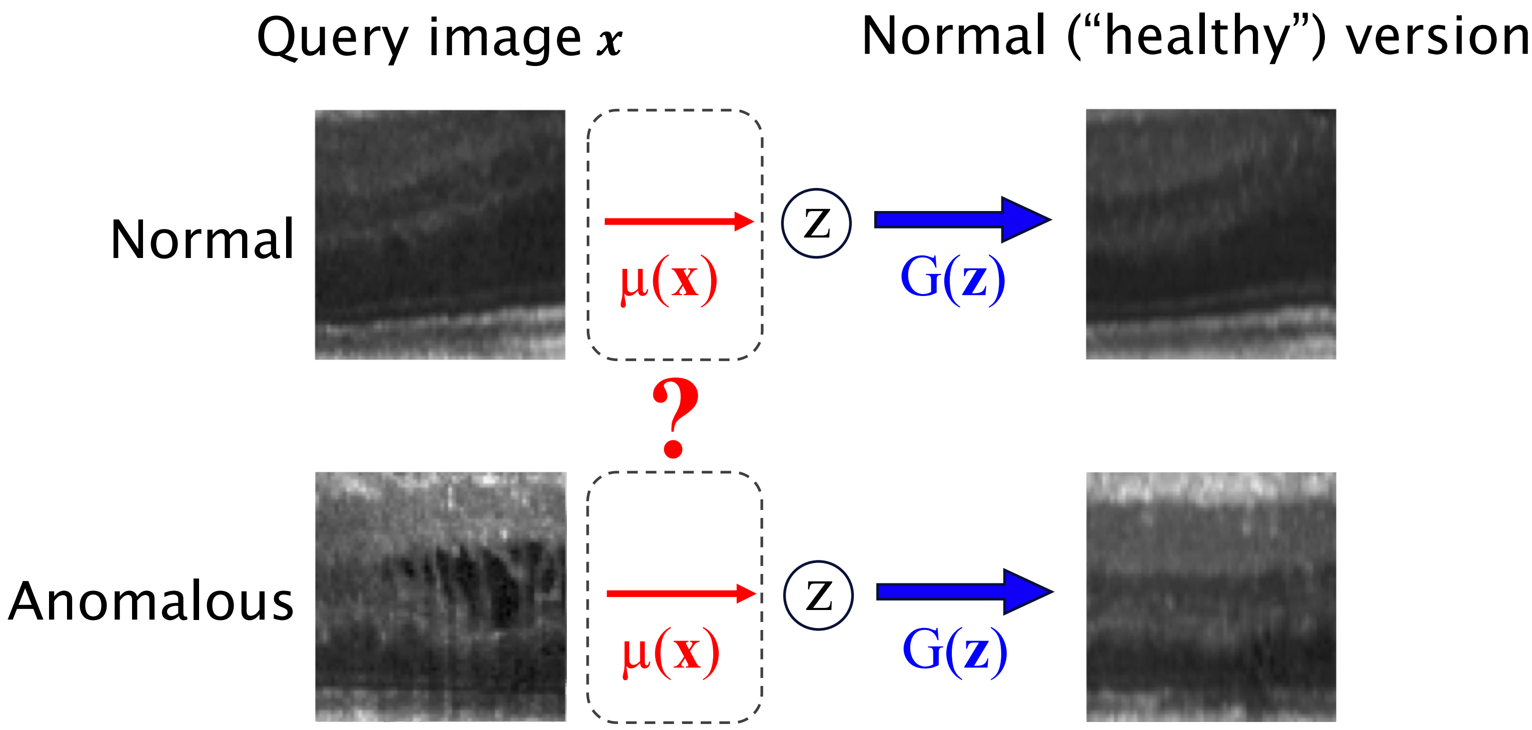
   Generating the closest “normal” image to a given query image via mapping to the latent space

   - *Mapping loss function*

2. **Anomaly detection:**
   - Detection of anomalous images (*Anomaly scoring*)
   - Detection of anomalous regions within images (*Residual*)

# Anomaly scoring: Main idea

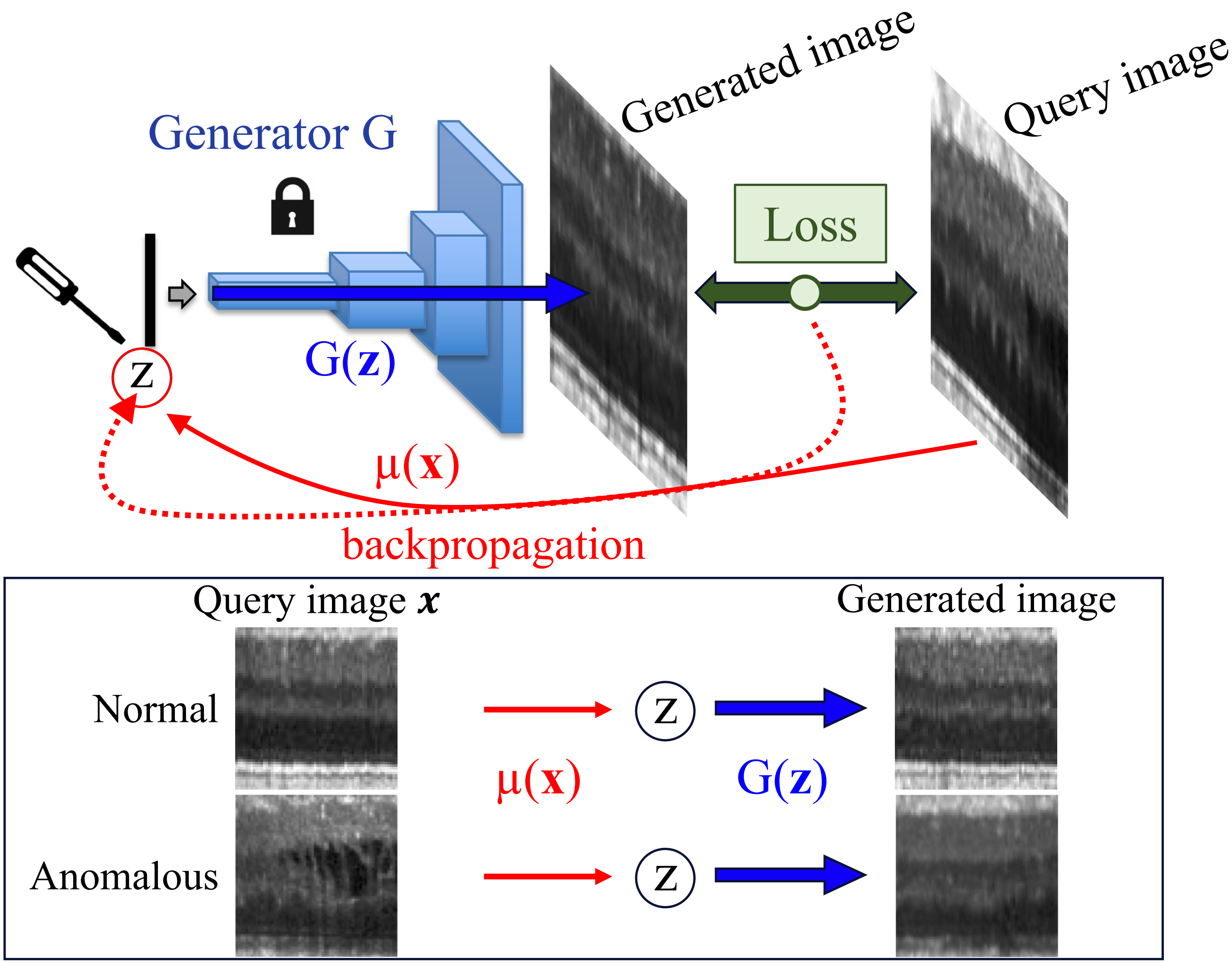


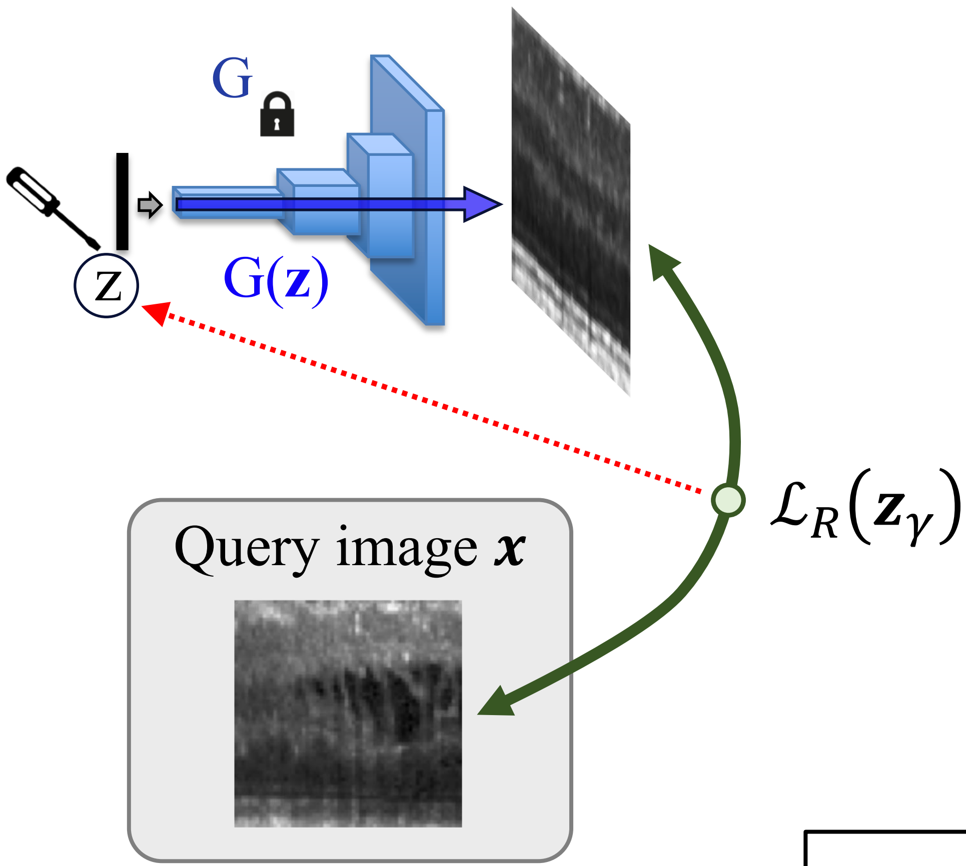
# Normality mapping techniques

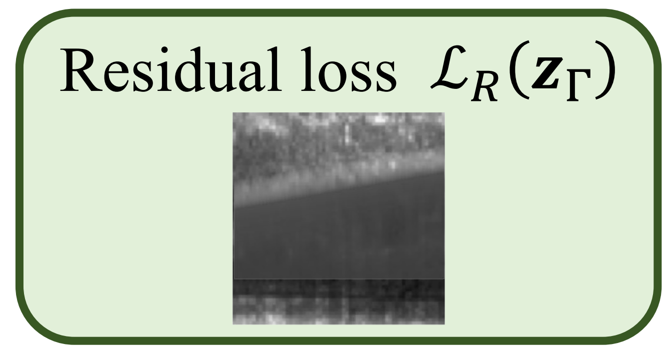
- AnoGAN
  - Normality mapping via a backpropagation
- f-AnoGAN
  - Normality mapping via a separate **encoder** training step

# AnoGAN *Normality* mapping



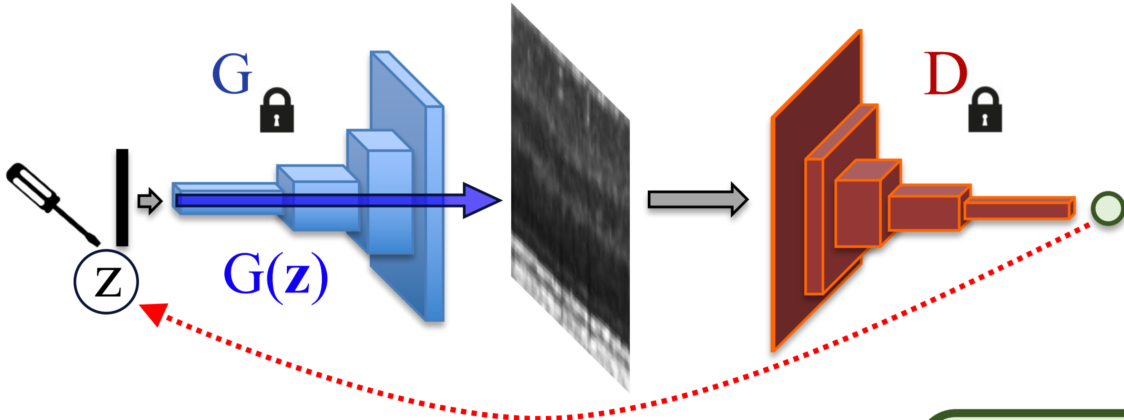
# AnoGAN *Normality* mapping: Ingredient 1



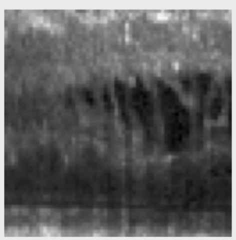

$$\boxed{\mathcal{L}_R(\boldsymbol{z}_\gamma) = \sum \left|\boldsymbol{x} - G(\boldsymbol{z}_\gamma)\right|}$$

# AnoGAN *Normality* mapping: Ingredient 2

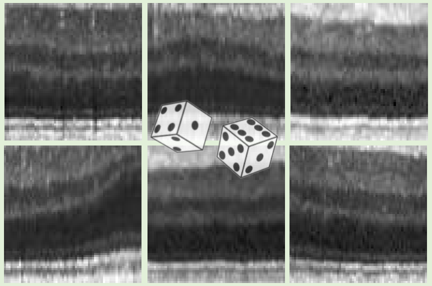



$\mathcal{L}_{\hat{D}}\left(\boldsymbol{z}_{\gamma}\right)$

Query image $\boldsymbol{x}$



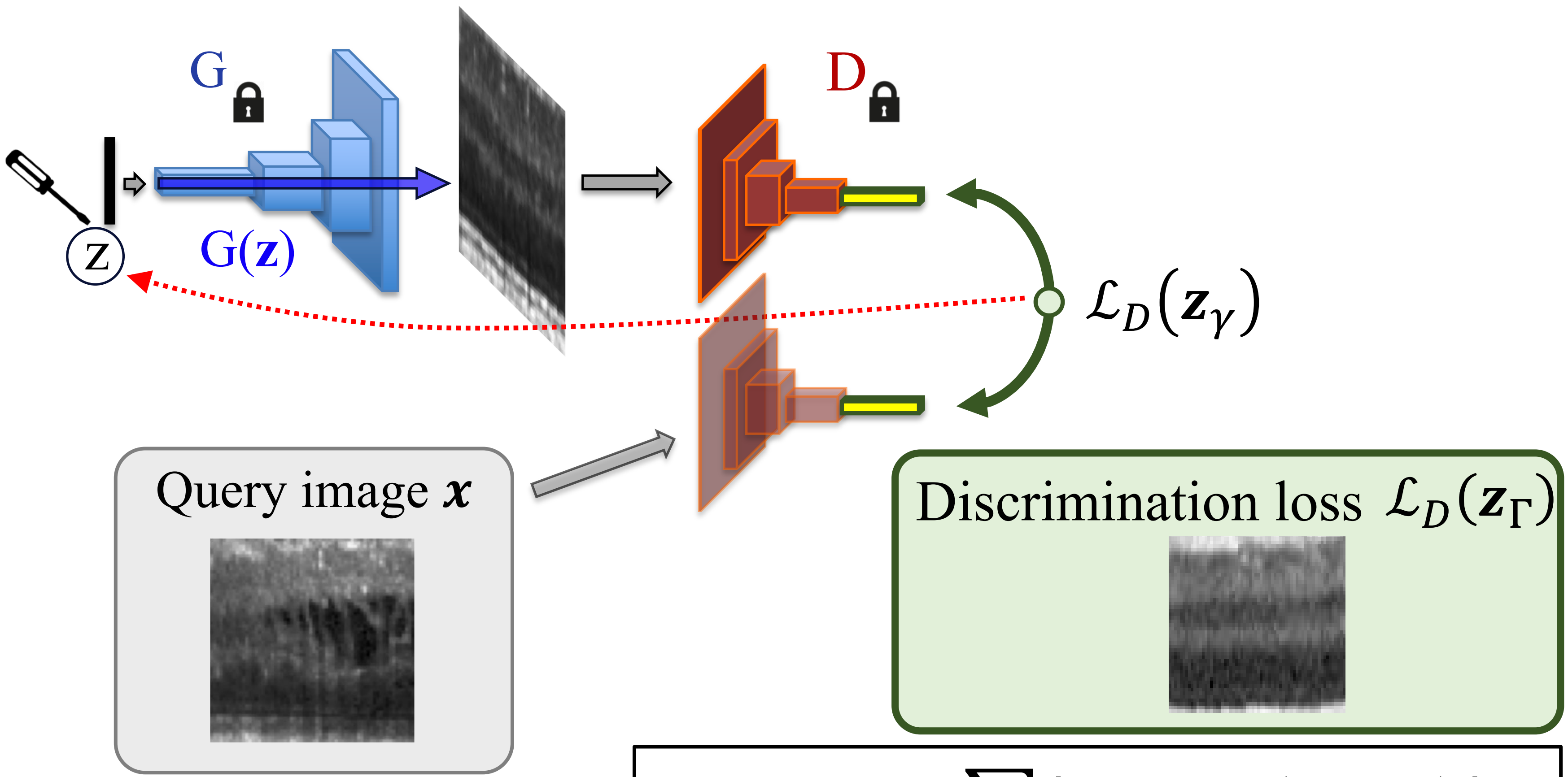
Discrimination loss $\mathcal{L}_{\hat{D}}(\boldsymbol{z}_{\Gamma})$



$$\mathcal{L}_{\hat{D}}\left(\boldsymbol{z}_{\gamma}\right) = -\log D\left(G\left(\boldsymbol{z}\right)\right)$$

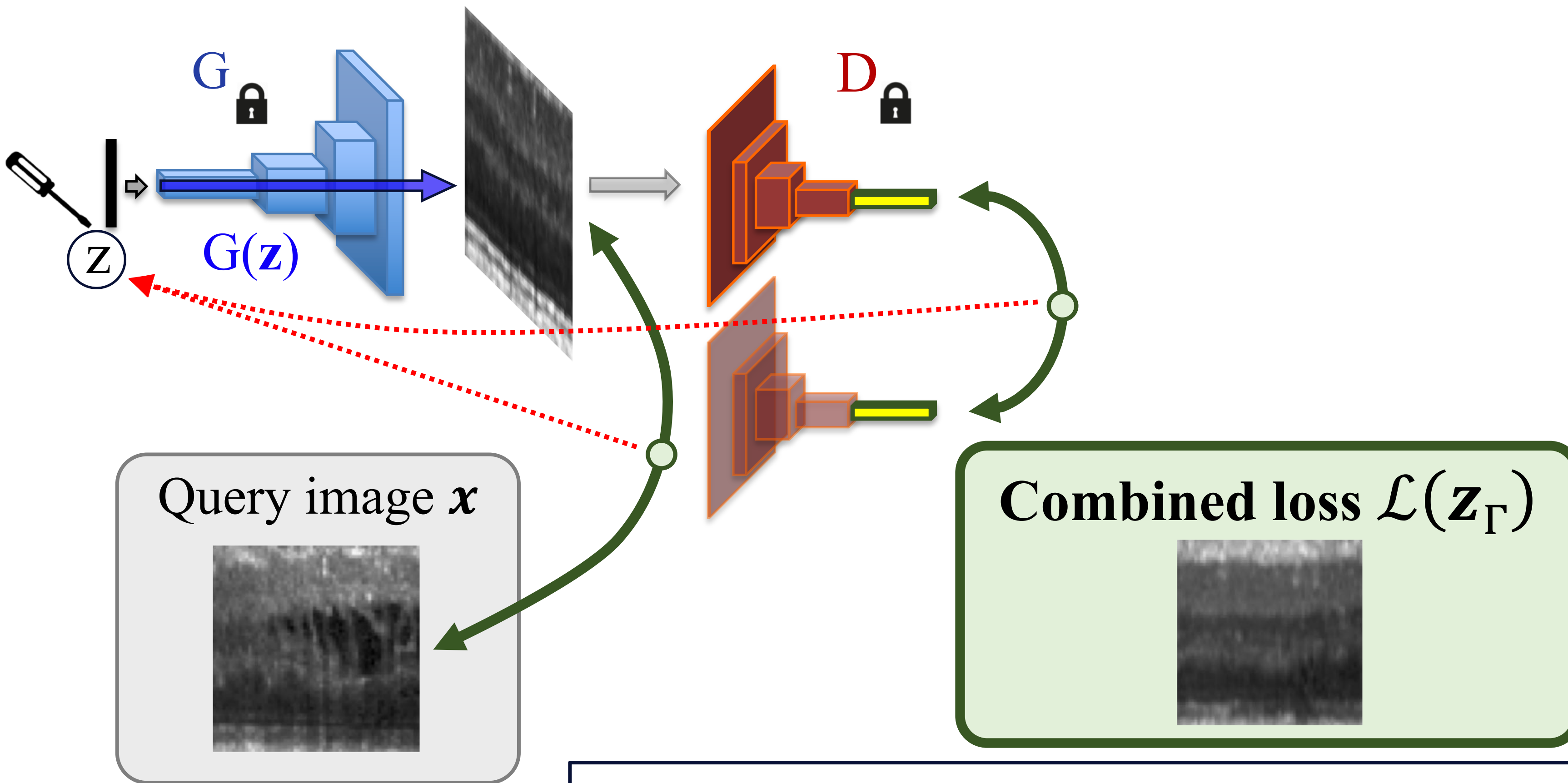
# **AnoGAN** *Normality* mapping: Ingredient 2 (revised)

*Feature matching* [Salimans et al., 2016]




$$\mathcal{L}_D(\boldsymbol{z}_\gamma) = \sum \left| \mathrm{f}(\boldsymbol{x}) - \mathrm{f}\left(G(\boldsymbol{z}_\gamma)\right) \right|$$

# **AnoGAN** *Normality* mapping: Combined loss function



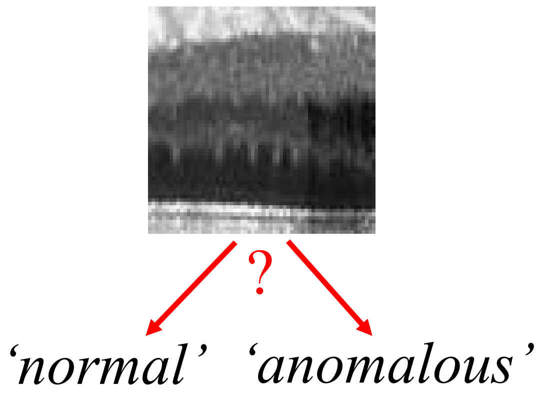

$$\mathcal{L}(\boldsymbol{z}_\gamma) = (1-\lambda)\cdot\mathcal{L}_R(\boldsymbol{z}_\gamma) + \lambda\cdot\mathcal{L}_D(\boldsymbol{z}_\gamma)$$

# **AnoGAN** Anomaly detection

1. **Anomaly score** $A(\boldsymbol{x})$:
   Detection of anomalous images

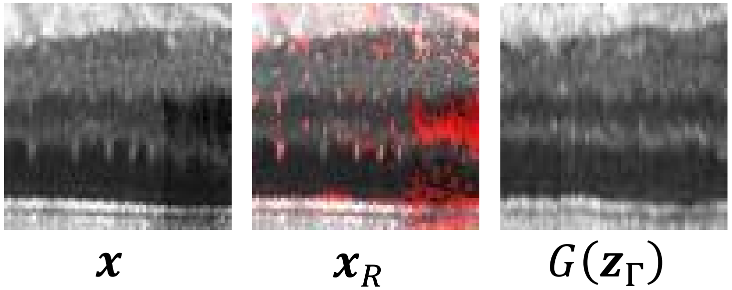
$$\boxed{A(\boldsymbol{x}) = (1-\lambda)\cdot R(\boldsymbol{x}) + \lambda \cdot D(\boldsymbol{x}),}$$

- *residual score* $\mathrm{R}(\boldsymbol{x}) = \mathcal{L}_R(\boldsymbol{z}_\Gamma)$ and
- *discrimination score* $\mathrm{D}(\boldsymbol{x}) = \mathcal{L}_D(\boldsymbol{z}_\Gamma)$




2. **Residual image**:
   Detection of anomalous regions within images

$$\boxed{\boldsymbol{x}_R = |\boldsymbol{x} - G(\boldsymbol{z}_\Gamma)|}$$



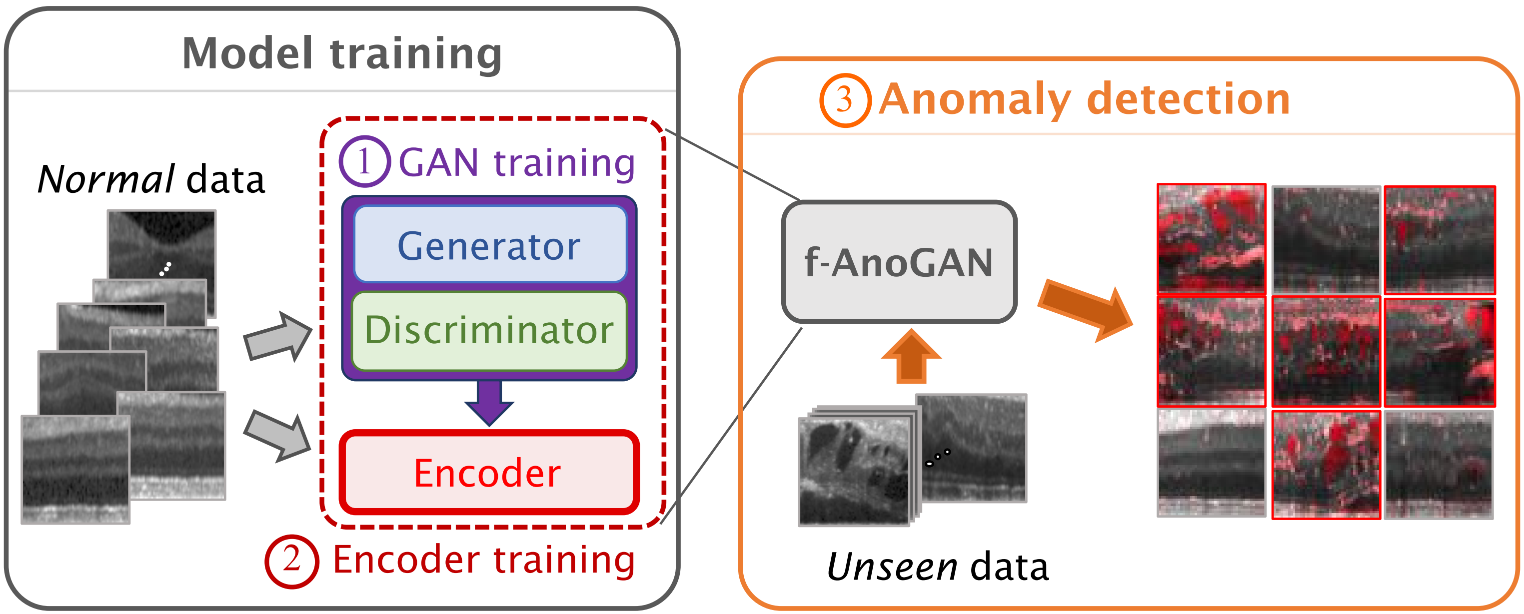
# Motivation: f-AnoGAN

- *AnoGAN*: runtimes limit the actual applicability
  - GAN training: ~20h
  - Processing full OCT volume: minutes - hours
- **Aim: f-AnoGAN:**
  - Speeding up inference → encoder training

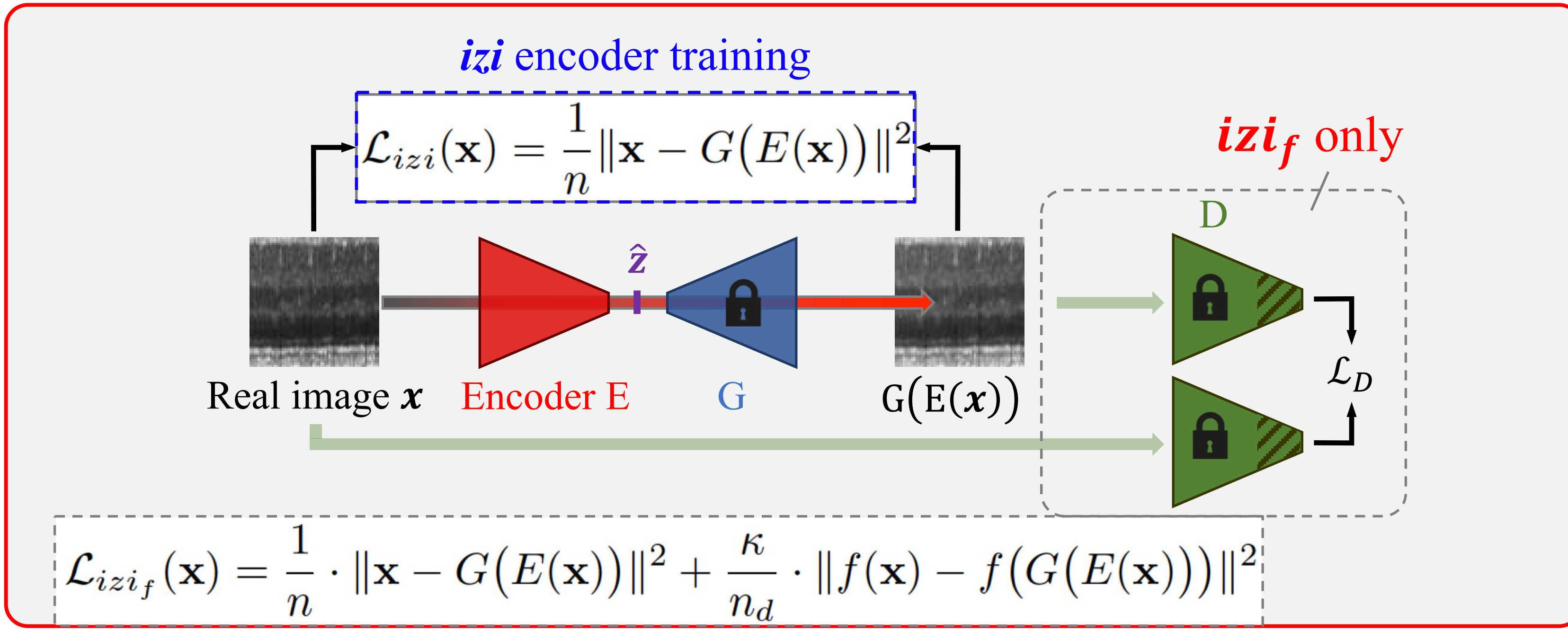
# f-AnoGAN

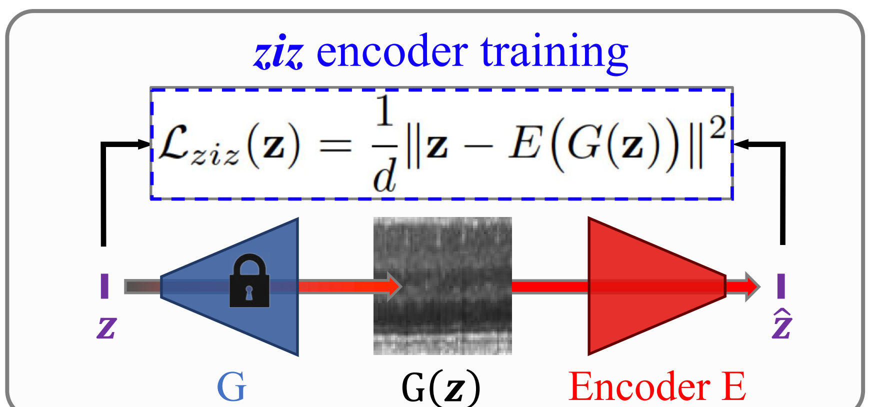
## Anomaly detection with GANs

# f-AnoGAN: Anomaly detection framework



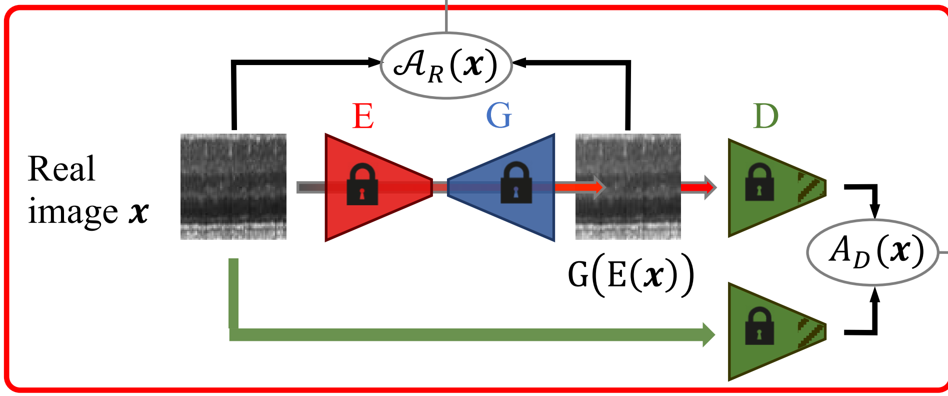
# f-AnoGAN *Normality* mapping



# f-AnoGAN Anomaly detection

$$\mathcal{A}_R(\mathbf{x}) = \frac{1}{n} \cdot \|\mathbf{x} - G(E(\mathbf{x}))\|^2$$

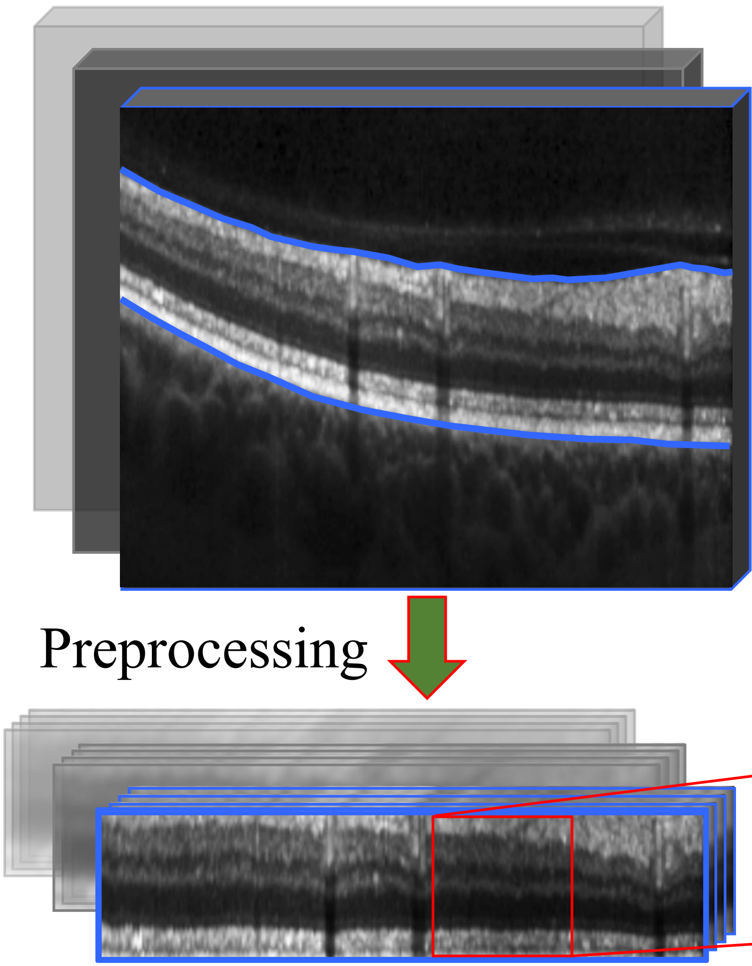



$$\mathcal{A}_D(\mathbf{x}) = \frac{1}{n_d} \cdot \|f(\mathbf{x}) - f(G(E(\mathbf{x})))\|^2$$

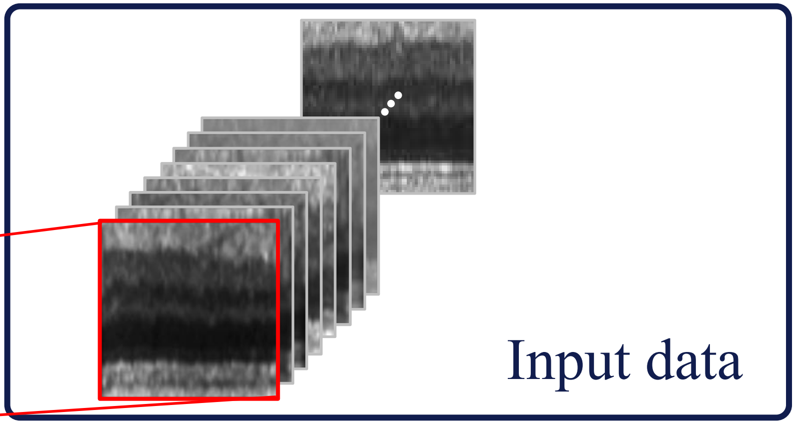
$$\boxed{\mathcal{A}(\mathbf{x}) = \mathcal{A}_R(\mathbf{x}) + \kappa \cdot \mathcal{A}_D(\mathbf{x})}$$

$$\boxed{\dot{\mathcal{A}}_R(\mathbf{x}) = |\mathbf{x} - G(E(\mathbf{x}))|}$$

# Experiments

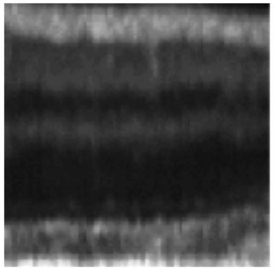


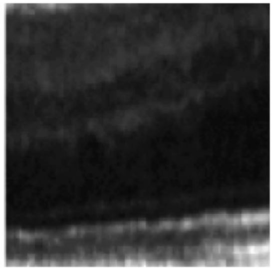

- Spectral Domain Optical Coherence Tomography
- OCT scan: $496 \times 512 \times 49$
- Image patches: $64 \times 64\, px$



# Can the model generate realistic images?

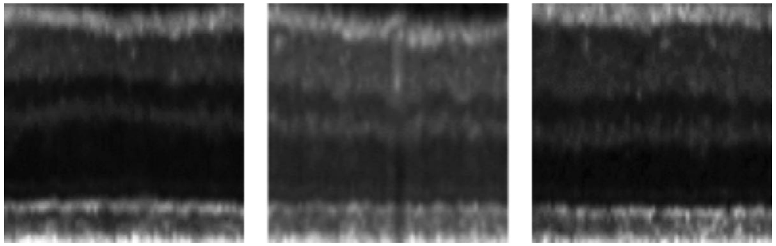
- **"Visual Turing test"** (cf. Chuquicusma et al. (2018))

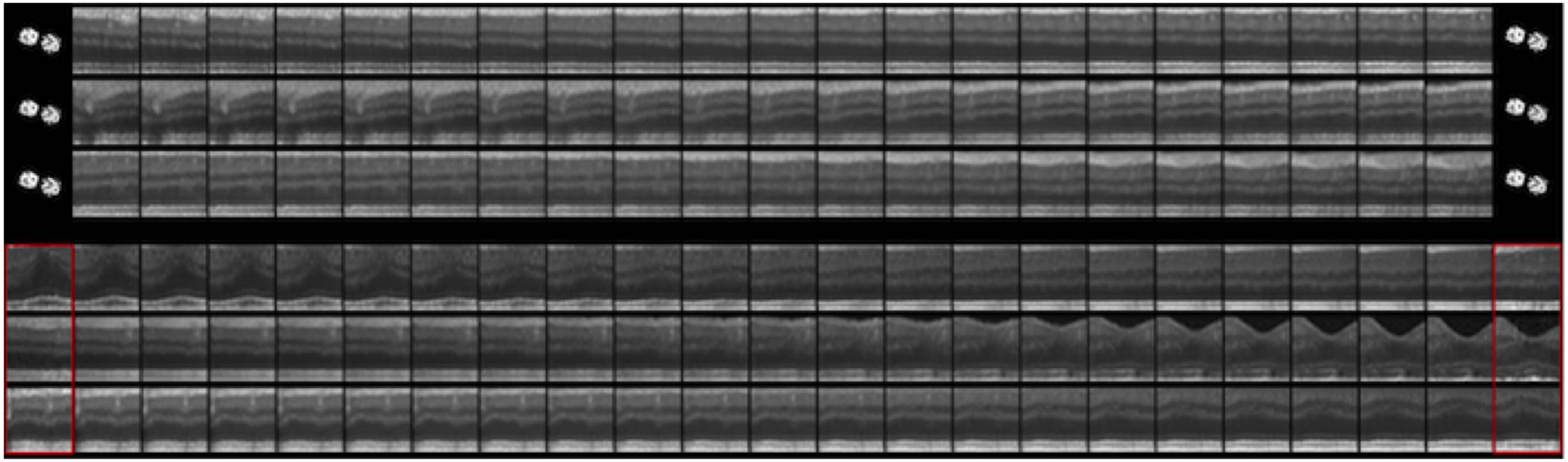




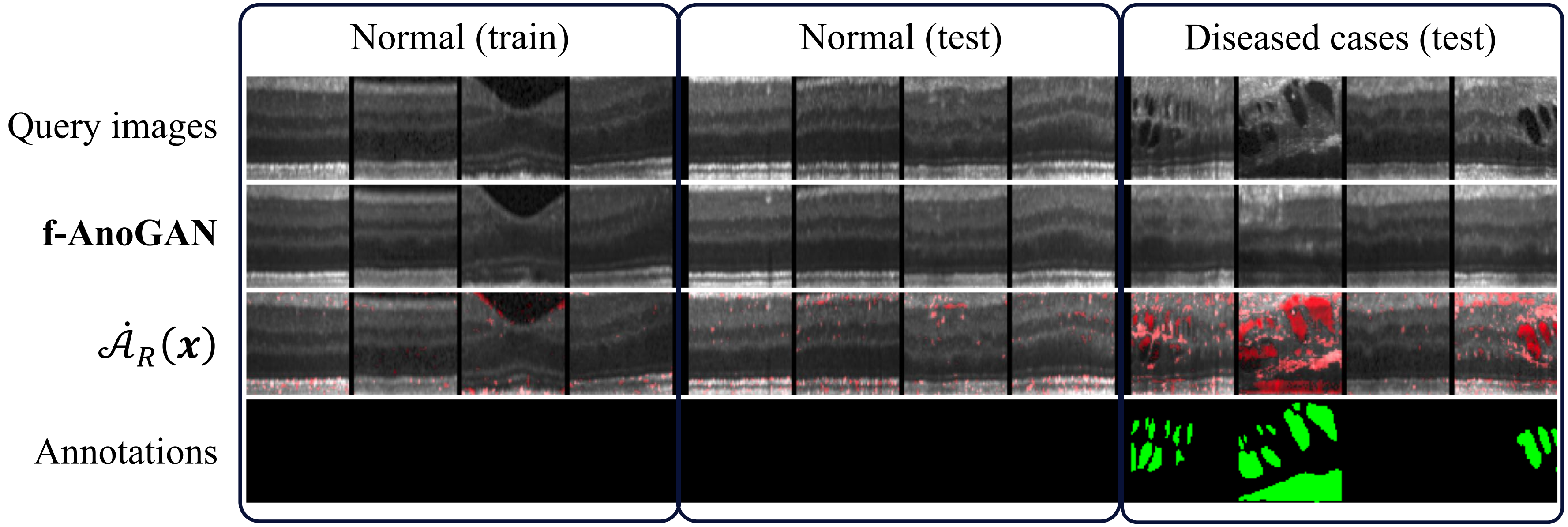
- 100 images (50 real, 50 generated)
- Image size: 64x64 pixels
- 2 retina experts: mean accuracy on telling generated from real normal images apart: 0.44
- Consensus between clinical experts: 0.58

# Does the model capture normal variability?



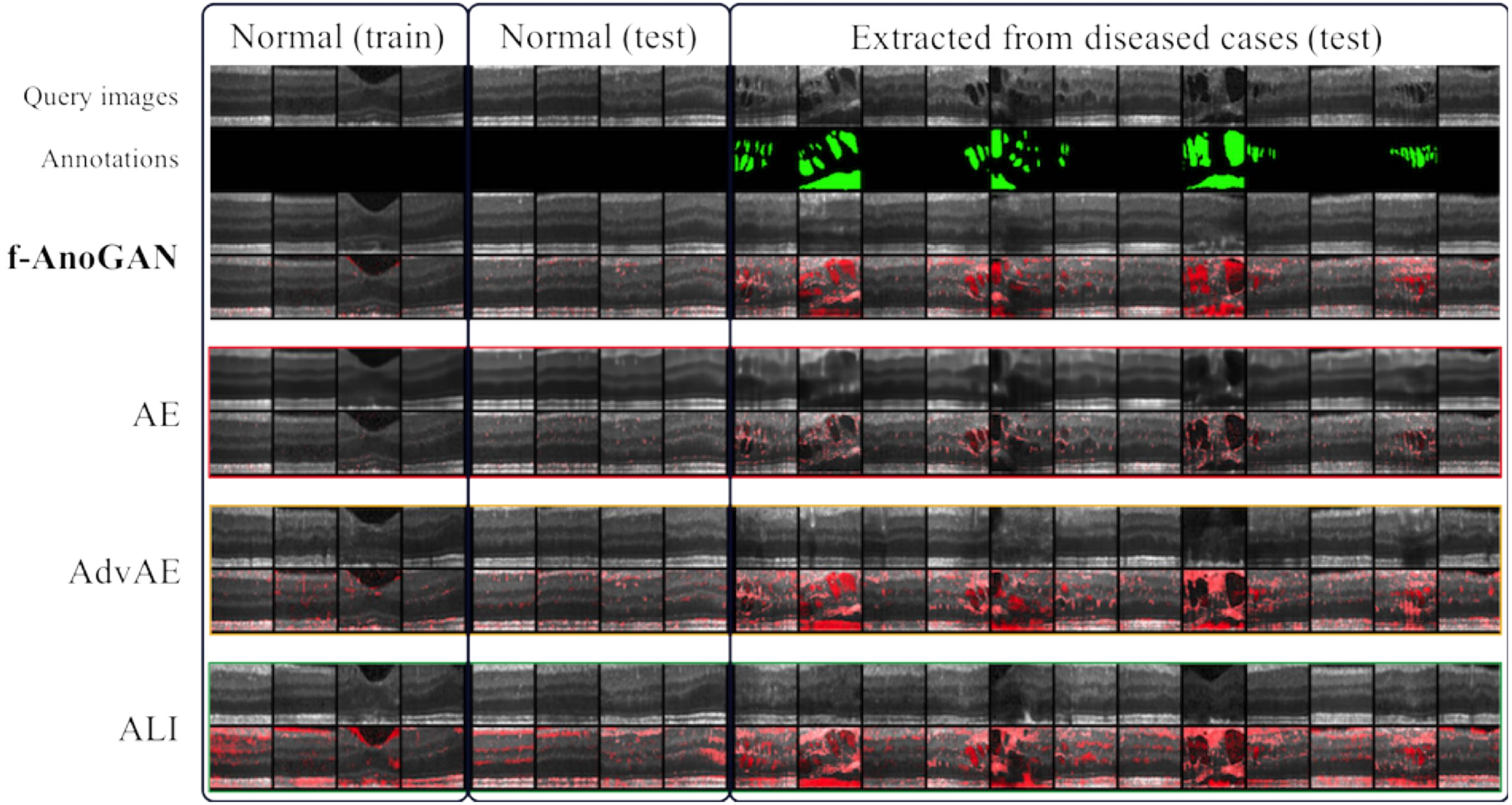


# Can the model localize anomalies in images?



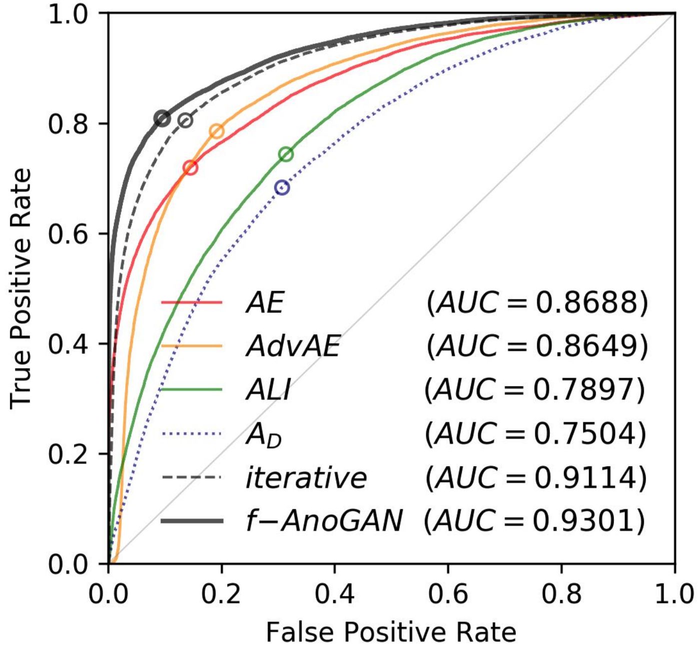

$$\boxed{\dot{A}_R(\mathbf{x}) = |\mathbf{x} - G(E(\mathbf{x}))|}$$

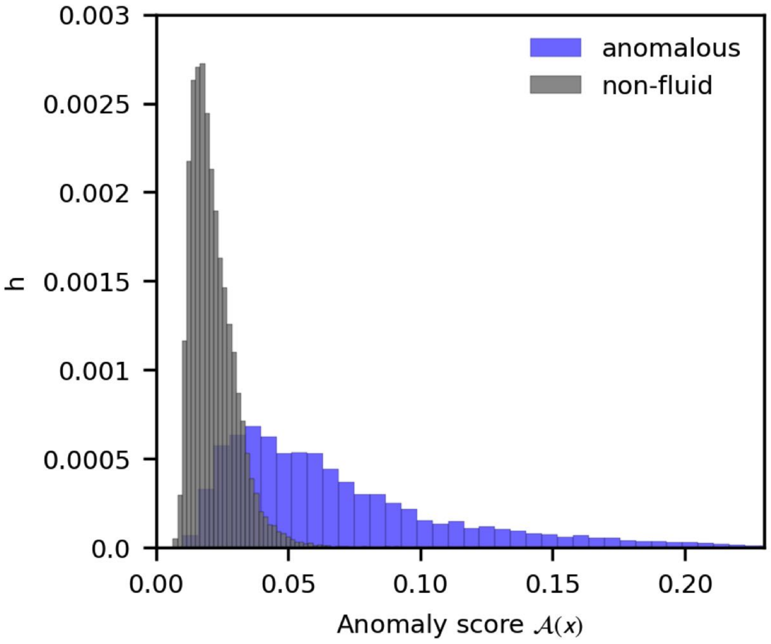
# 2. Can the model localize anomalies in images?




$$\boxed{\dot{A}_R(\mathbf{x}) = |\mathbf{x} - G(E(\mathbf{x}))|}$$

# 3. Does the model detect anomalous images accurately?




| | Precision | Sensitivity | Specificity | f-score | AUC | 70,000 test images |
|---|---|---|---|---|---|---|
| AE | 0.6824 | 0.7195 | 0.8550 | 0.7005 | 0.8688 | |
| AdvAE | 0.6405 | 0.7856 | 0.8092 | 0.7057 | 0.8649 | |
| ALI | 0.5063 | 0.7434 | 0.6863 | 0.6023 | 0.7897 | |
| $A_D$ | 0.4909 | 0.6831 | 0.6931 | 0.5713 | 0.7504 | |
| iterative | 0.7202 | 0.8049 | 0.8645 | 0.7602 | 0.9114 | 2 days (300 updates) |
| *f-AnoGAN* | **0.7863** | **0.8091** | **0.9049** | **0.7975** | **0.9301** | 20 sec. |

# Conclusion

# Conclusion

- (f-)AnoGAN:
  only requires a-priori definition (based on volume-level grading) and collection of ”normal” data
- Approach is able to detect anomalies that are not present in the training data
- Capable to discover novel anomalies
- Results demonstrate good performance on the image-level detection of known anomalies and the capability to segment those lesions

# Conclusion

**Limitations**

1. In some clinical settings, cases that do not contain any lesions are possibly only available to a certain extent
2. *Impurity* of “healthy” data => incorporated as normal variability of healthy data

# f-AnoGAN code on Github

tSchlegl/f-AnoGAN

www.linkedin.com