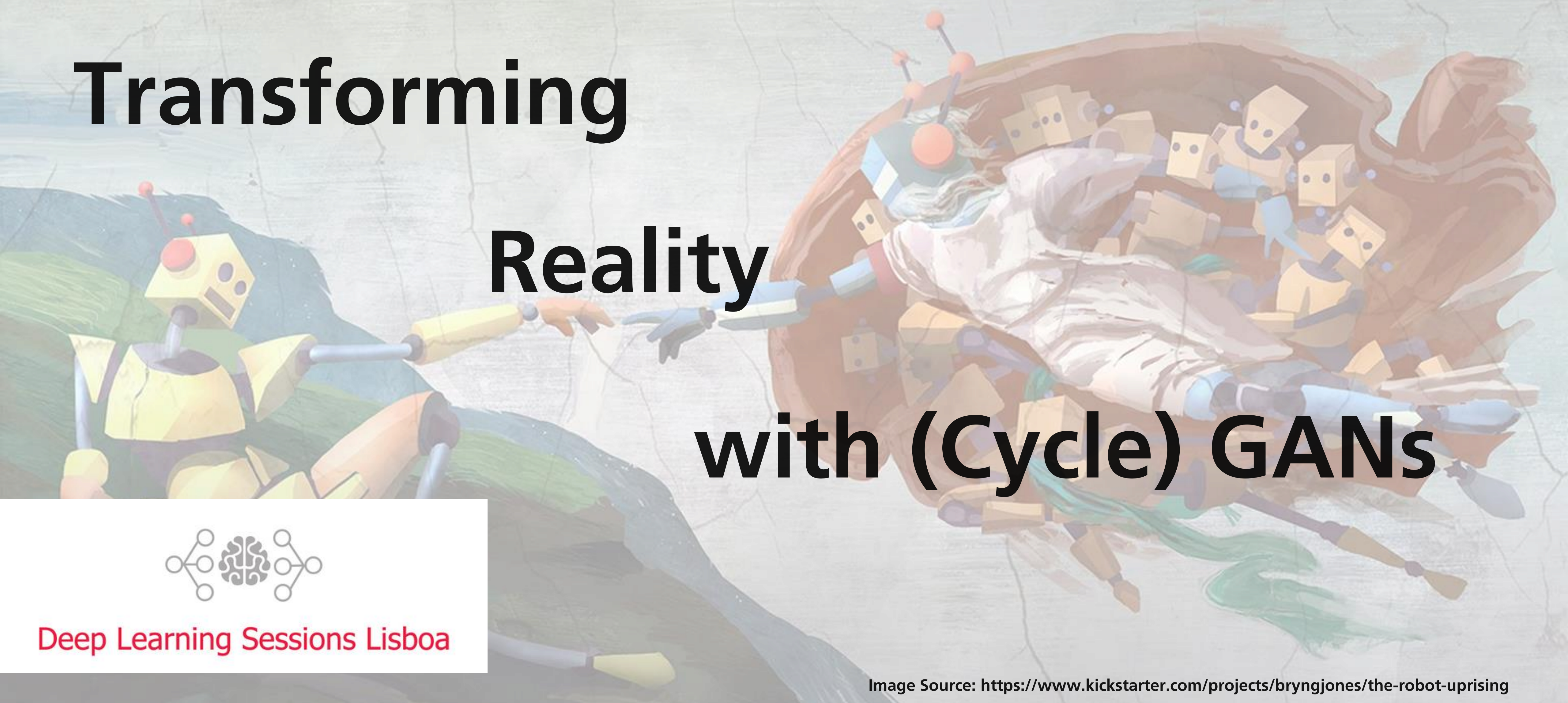
Transforming Reality with (Cycle) GANs



Deep Learning Sessions Lisboa

Image Source: https://www.kickstarter.com/projects/bryngjones/the-robot-uprising

Fraunhofer PORTUGAL

Transforming Reality with (Cycle) GANs

- A brief introduction…
- Generative Adversarial Networks – a Recap
- Image Translation
  - CycleGAN & Friends
- GANs at Work @ Fraunhofer Portugal
  - Skin lesions
  - Retinal images
- A quick tutorial

Fraunhofer PORTUGAL

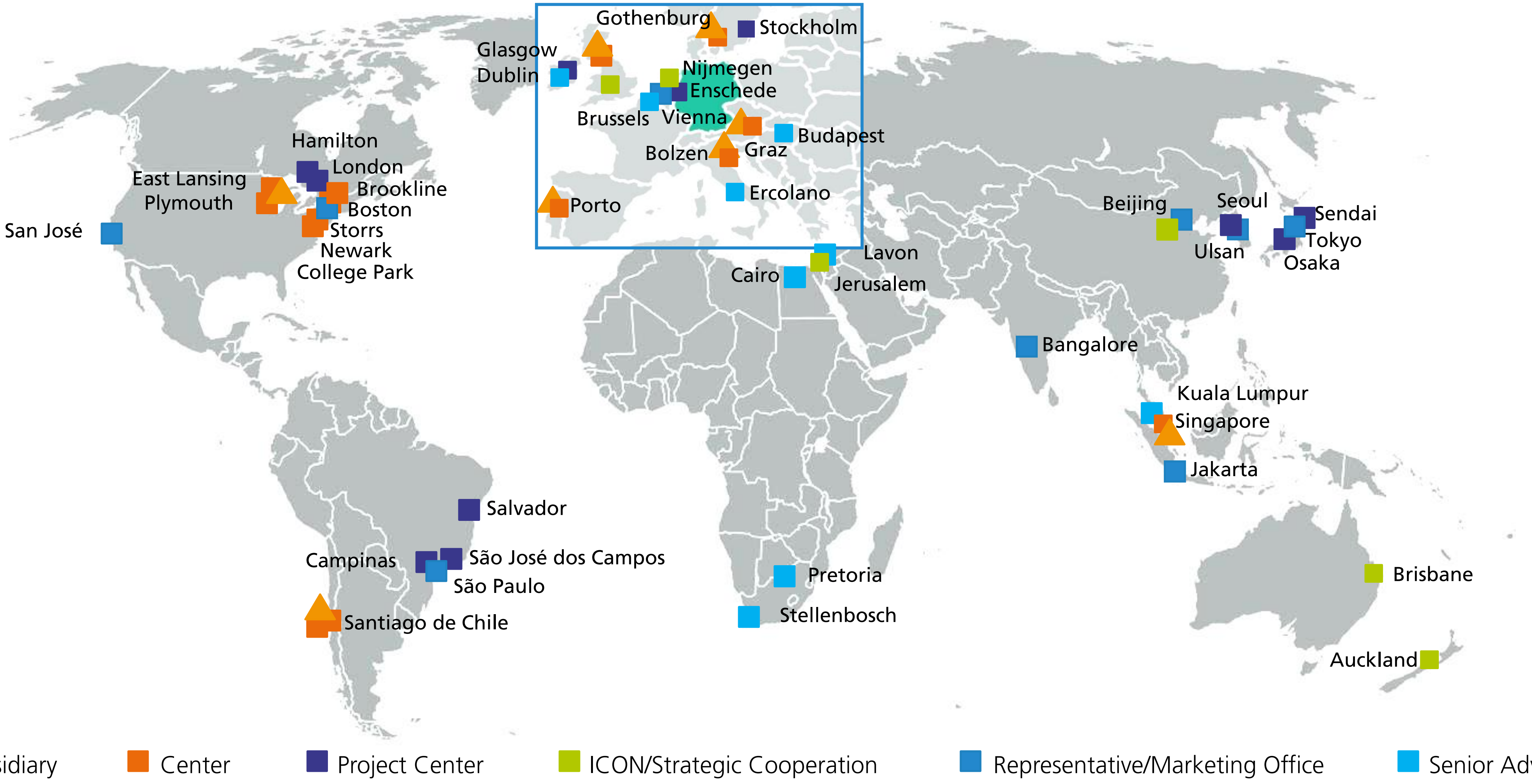
A brief introduction…

André V. Carreiro

- Senior Scientist at the Intelligent Systems Group of Fraunhofer Portugal AICOS
  - Computer Vision and Multimodal AI
  - Deep Learning
- PhD in Biomedical Eng. – IS Técnico (Univ. Lisbon)
  - Integrative Data Mining applied to Neurodegenerative diseases (ALS)
- Experience at 2 startups
  - Data science full-stack – from Database to ML layers and applications
  - Deep Learning for Human Behavior Analysis from Video

Fraunhofer Portugal

Fraunhofer – Gesellschaft



Fraunhofer PORTUGAL

Fraunhofer Portugal

Fraunhofer Operational Model and Technology Transfer

BASE FUNDING

~1/3 Revenues

DIRECT INDUSTRY CONTRACTS

~1/3 Revenues

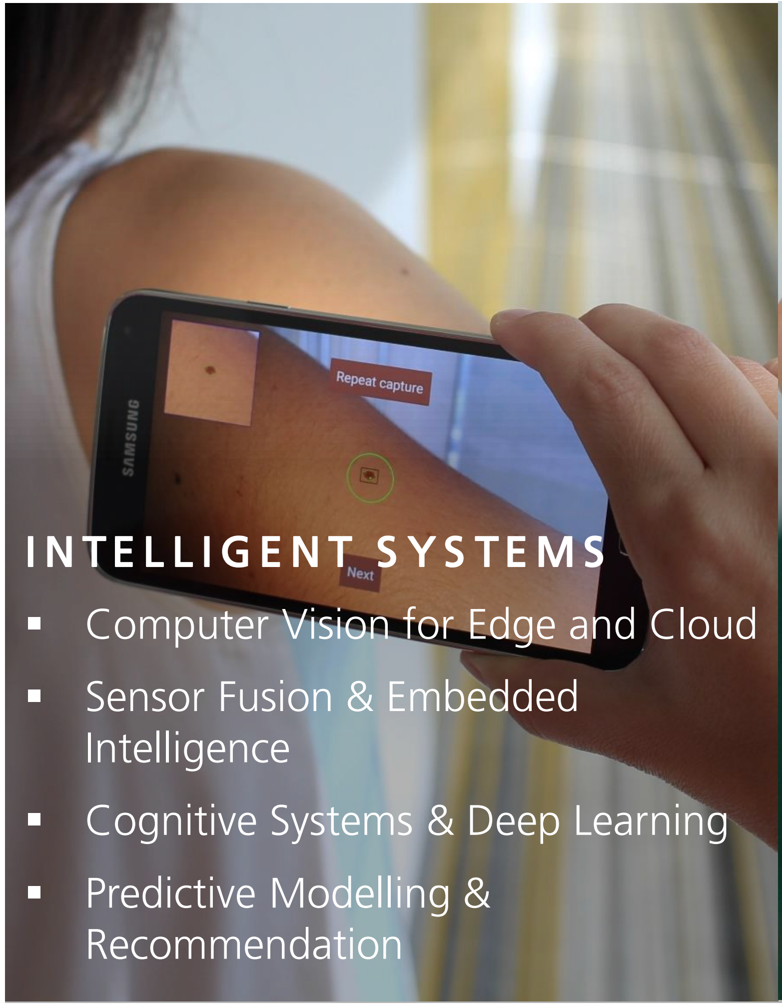
NATIONAL / INTERNATIONAL R&D PROGRAMMES

~1/3 Revenues

Fraunhofer PORTUGAL

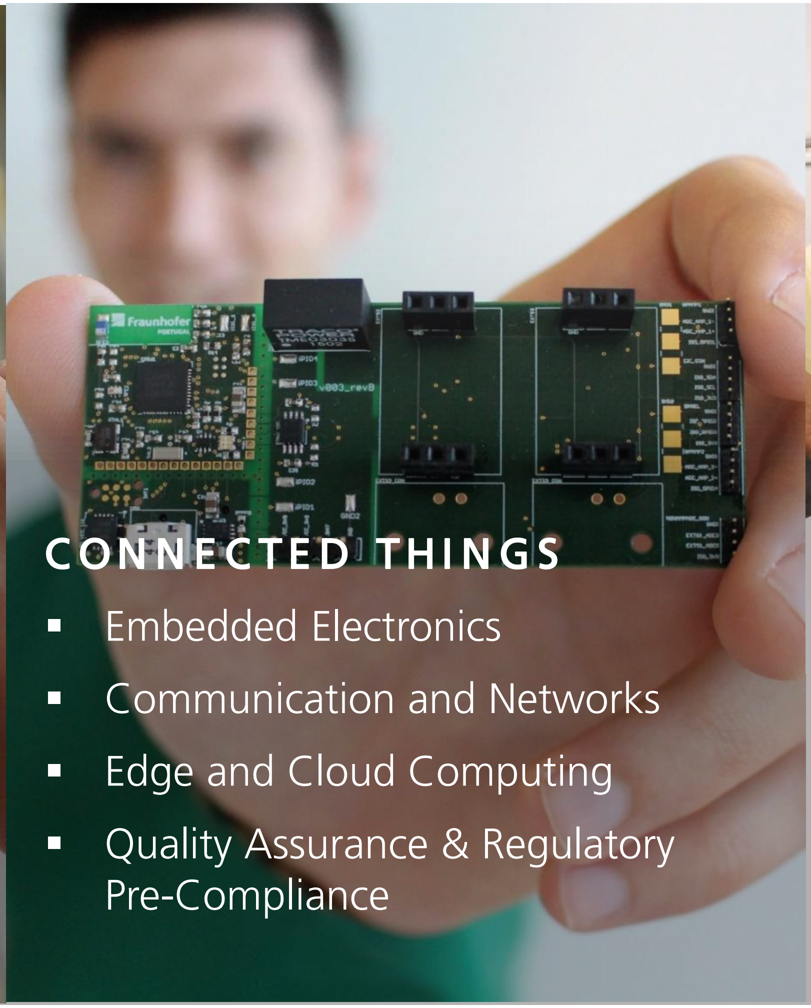
Fraunhofer AICOS

Purpose and Scientific Areas



INTELLIGENT SYSTEMS

- Computer Vision for Edge and Cloud
- Sensor Fusion & Embedded Intelligence
- Cognitive Systems & Deep Learning
- Predictive Modelling & Recommendation



CONNECTED THINGS

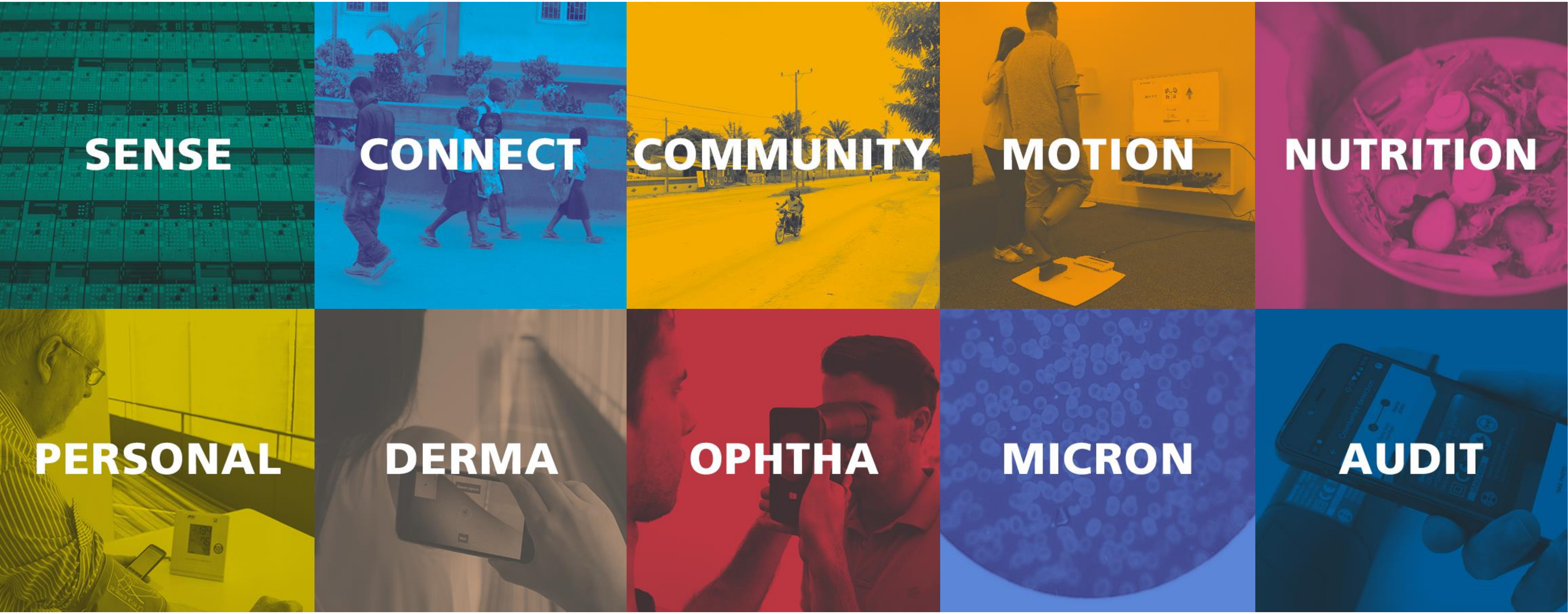
- Embedded Electronics
- Communication and Networks
- Edge and Cloud Computing
- Quality Assurance & Regulatory Pre-Compliance

HUMAN-CENTRED DESIGN

- Understanding people in diverse settings
- Co-designing meaningful technologies
- Technology assessment in real life

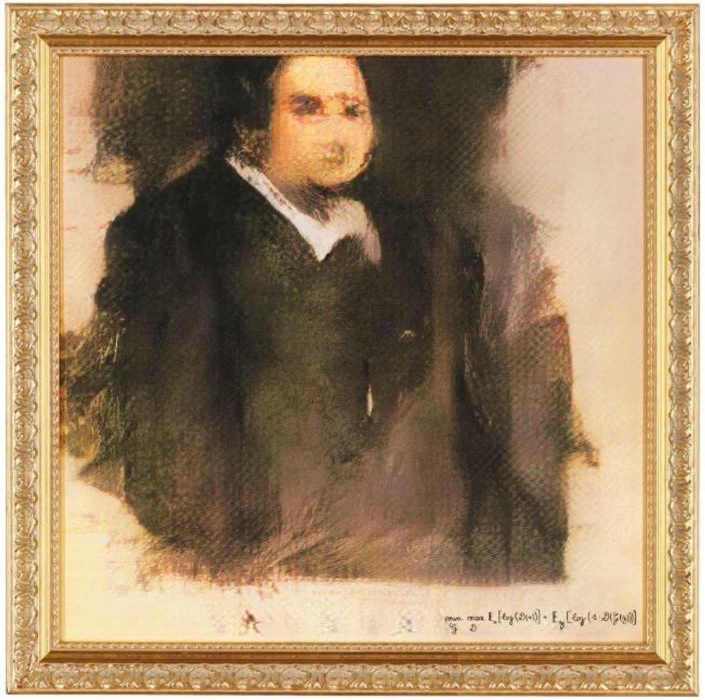
Fraunhofer AICOS

Portfolio



Generative Adversarial Networks – a Recap

- "Adversarial training is the coolest in ML, in the last 20 years"
  – Yann Lecun, 2016
- Introduced in 2014 by Ian Goodfellow
- As the name suggests, (at least) two networks are pit against each other
- Can learn to mimic any distribution of data:
  - speech, **images**, text, music…



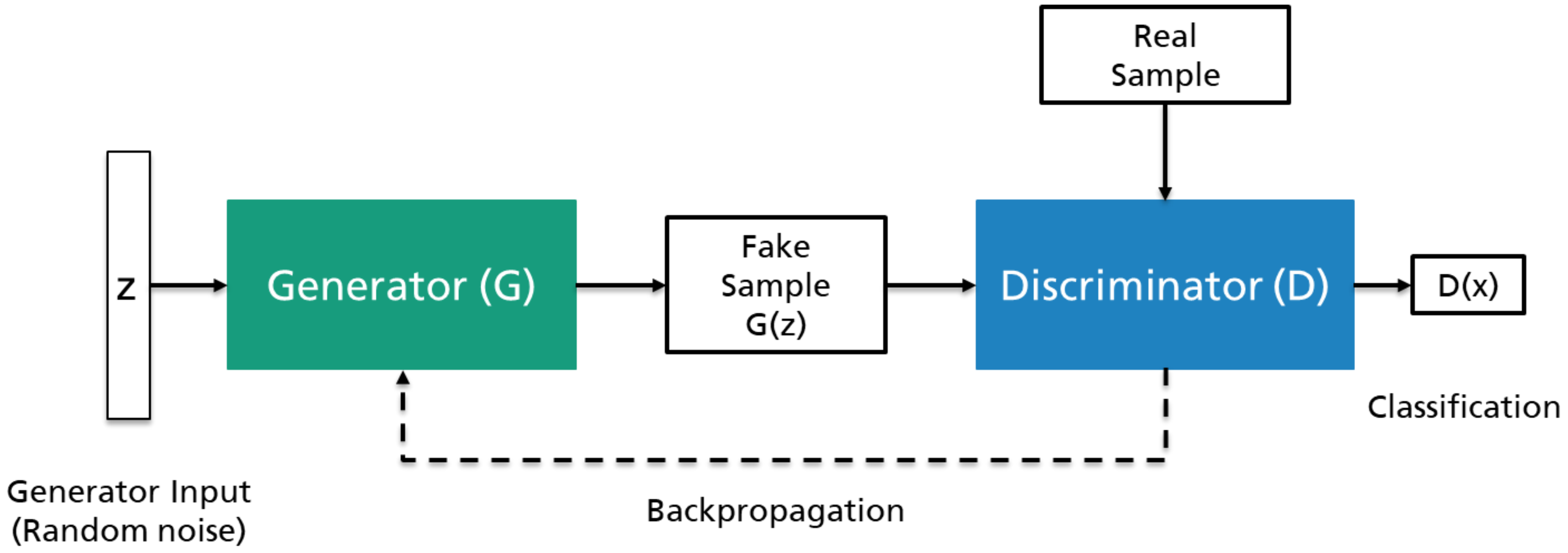
Portrait of Edmond de Belamy, a GAN-generated "painting", auctioned for 432 500$

Fraunhofer PORTUGAL

Generative Adversarial Networks – a Recap

- Generator vs. Discriminator, usually “playing” a zero-sum game (min-max)

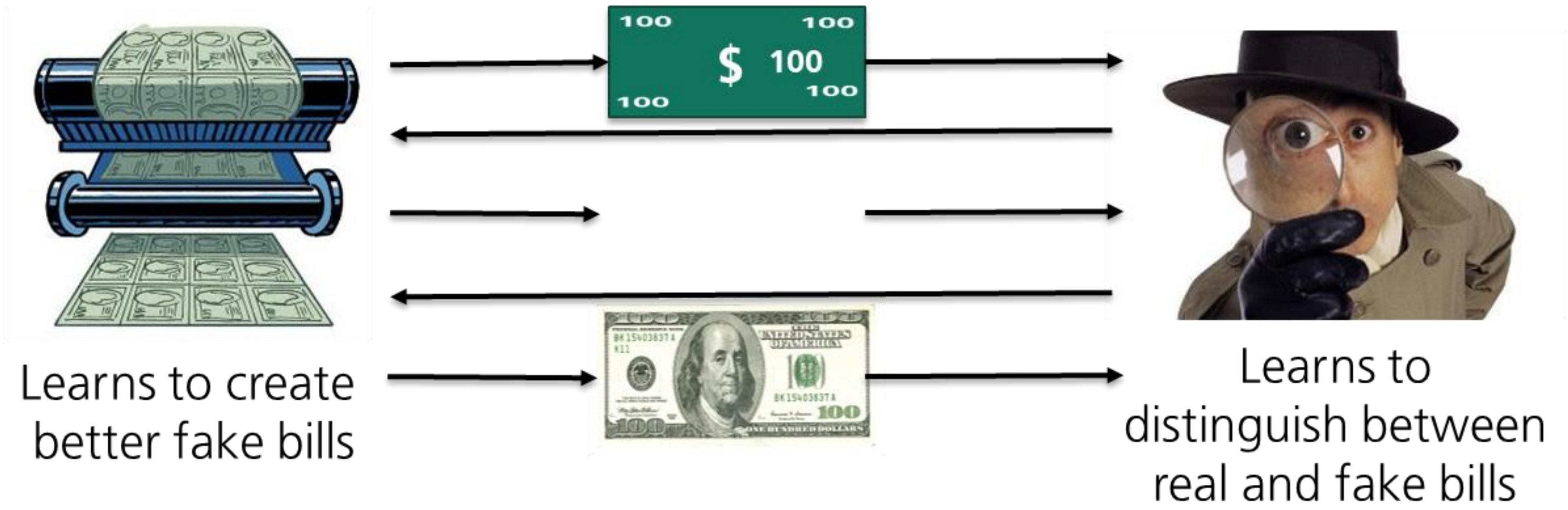
$$\min_G \max_D V(D,G) = \underbrace{\mathbb{E}_{x\sim p_{data}(x)}[\log D(x)]}_{\text{Maximize D accuracy in real}} + \underbrace{\mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]}_{\substack{\text{D: Maximize D accuracy in fakes} \\ \text{G: Minimize D accuracy (fool D)}}}$$



Fraunhofer PORTUGAL

Generative Adversarial Networks – a Recap

- Generator vs. Discriminator, usually “playing” a zero-sum game (min-max)



Fraunhofer PORTUGAL

Generative Adversarial Networks – a Recap

Generation

- GANs, as the name suggests, can learn to generate new samples from a learned data distribution

Fake Bedrooms: Radford, Metz & Chintala (2016)

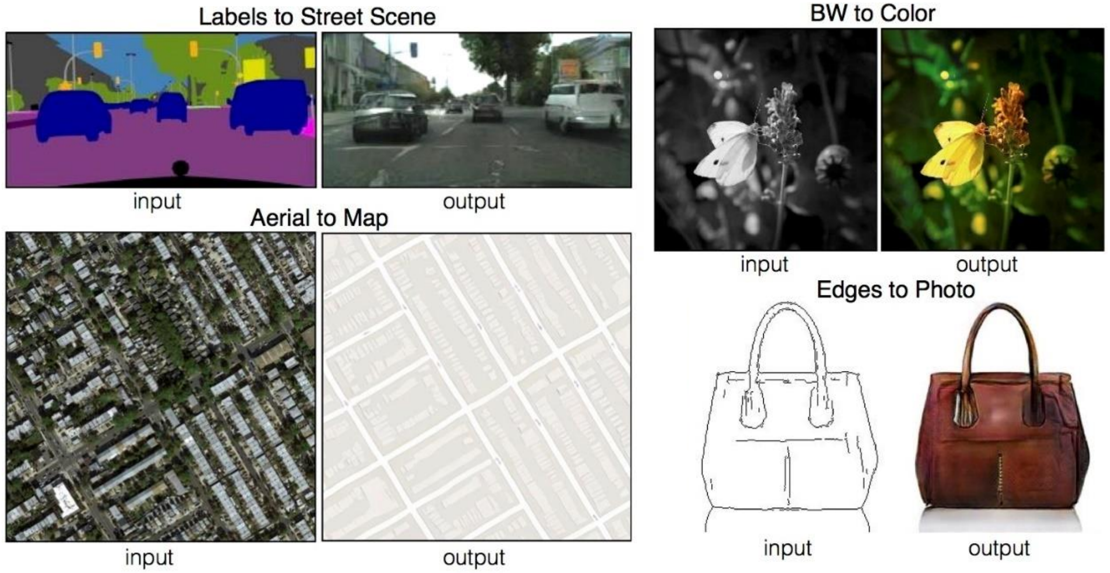
Progressive GAN: Karras *et al*. (2018)

Fraunhofer PORTUGAL

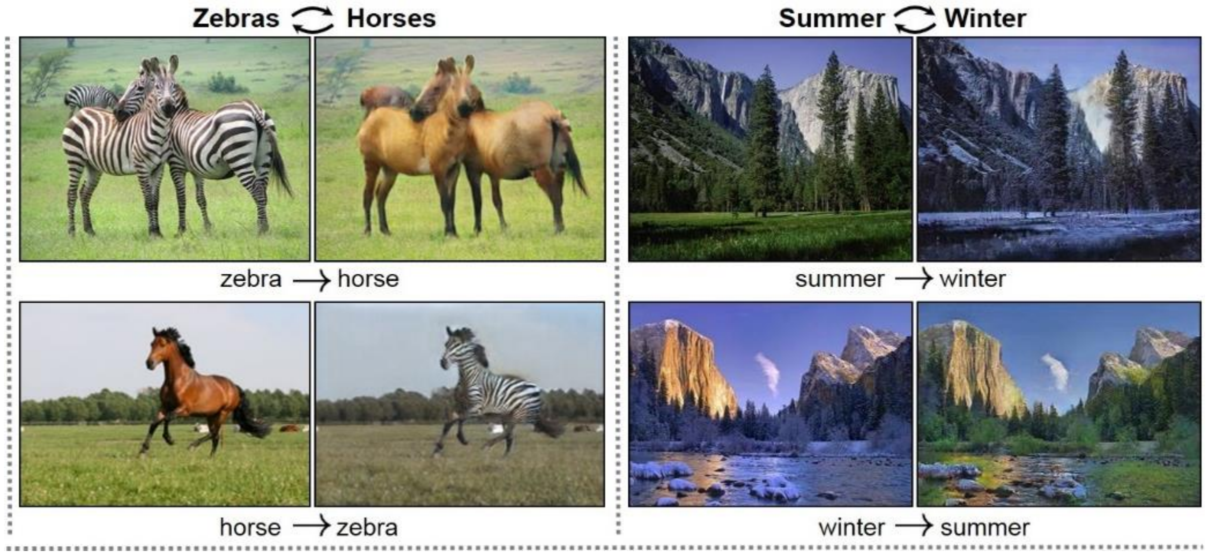
Generative Adversarial Networks – a Recap

Image Translation

- GANs can be used to "style" images and even change domains
- You can do it yourself! *Artbreeder.com*



Pix2Pix: Isola, Zhu, Zhou, and Efros (2017)



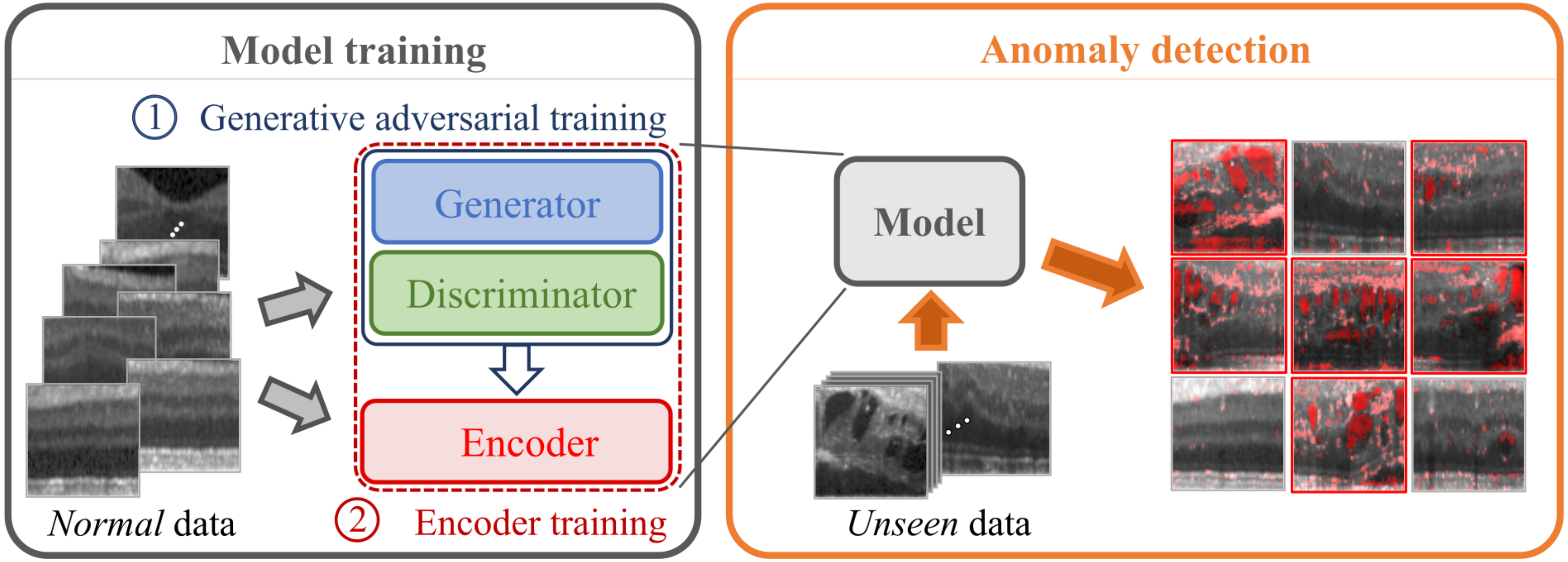
CycleGAN: Zhu, Park, Isola and Efros (2017)

Fraunhofer PORTUGAL

Generative Adversarial Networks – a Recap

Anomaly Detection

- Since they can learn virtually any distribution, you can use GANs also for Anomaly Detection



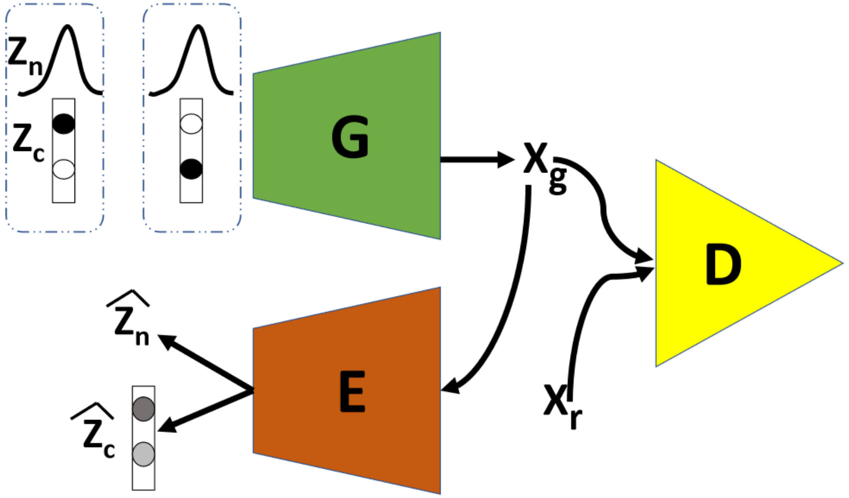
AnoGAN: Schlegl *et al* (2017)

Fast AnoGAN: Schlegl *et al* (2019)

Generative Adversarial Networks – a Recap

Clustering

- By learning a latent representation of the data, GANs can also learn to perform Clustering
  - ClusterGAN shows a particular advantage of allowing smooth interpolation in the latent space



ClusterGAN: Mukherjee *et al* (2018)

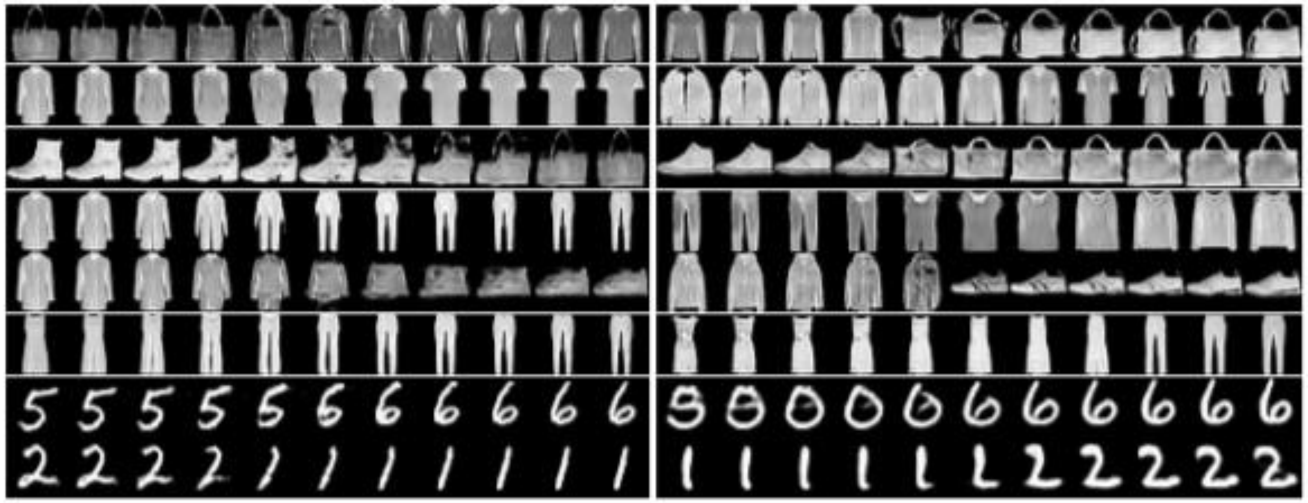


Figure 6: Comparison of Latent Space Interpolation : ClusterGAN (left) and vanilla WGAN (right)

Image Translation – Pix2Pix

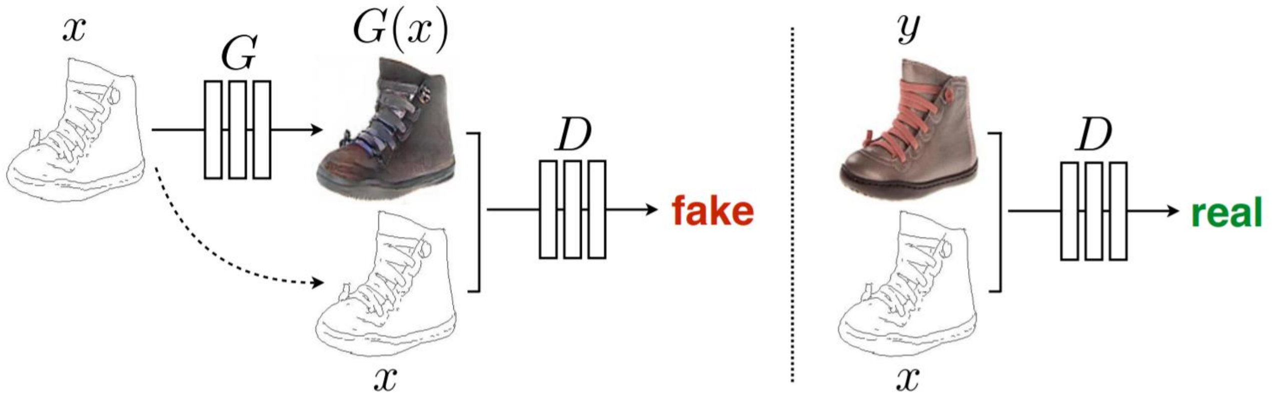
Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks.

- An example of a Conditional GAN (or cGAN)
  - Learns image-to-image mappings under a set of conditions
  - As opposed to Vanilla GANs, who generate images from a random distribution vector with no condition
  - **The drawback is that image mappings (x,y) are required**

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1]$$

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)$$



Fraunhofer PORTUGAL

Image Translation – Pix2Pix

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

- Also introduced the **PatchGAN** Discriminator
  - Instead of looking at the whole image to decide if it's Real or Fake, it looks at $N \times N$ patches
    - Runs convolutionally across the image, averaging all responses to provide the final output
  - Penalizes local structure and can be understood as a type of texture/style loss

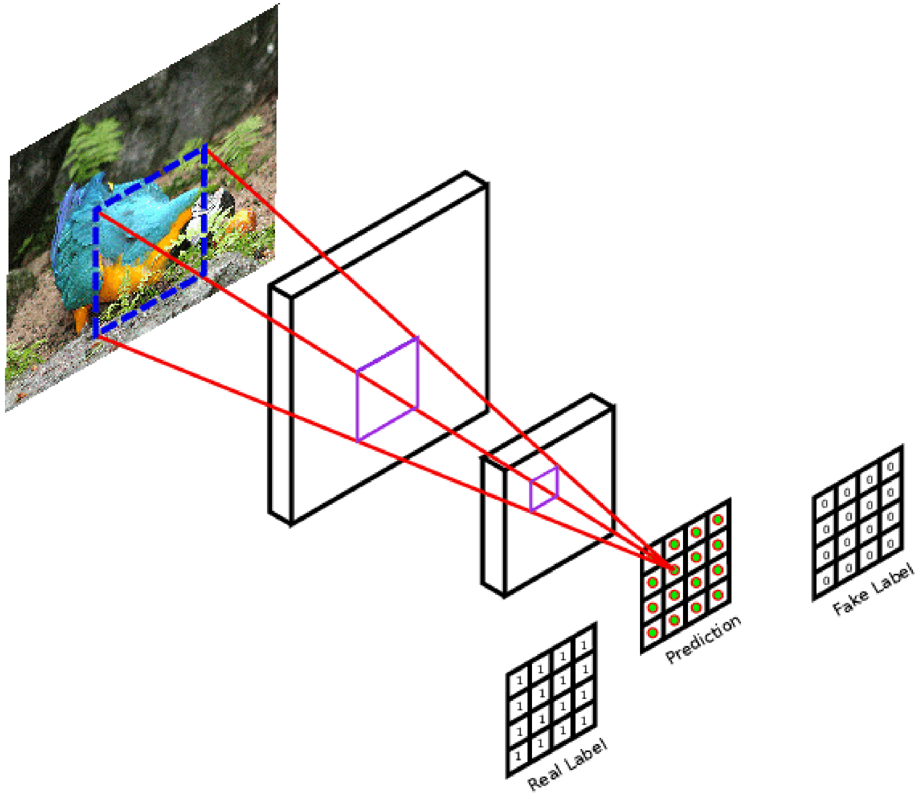
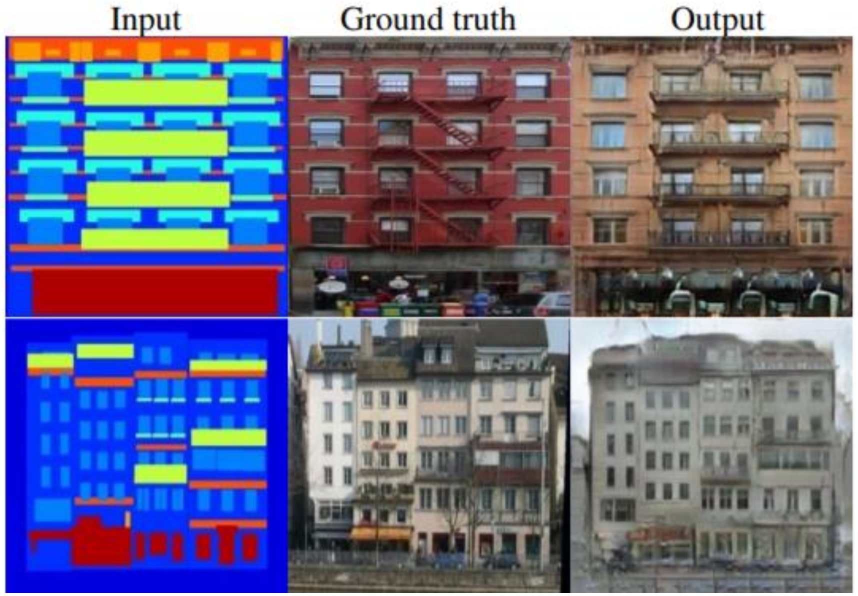


Image Translation – Pix2Pix

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



Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- Unlike with Pix2Pix, CycleGAN doesn't require image mapping pairs
  - **Unpaired Training**
  - Only needs sets of images for the two domains

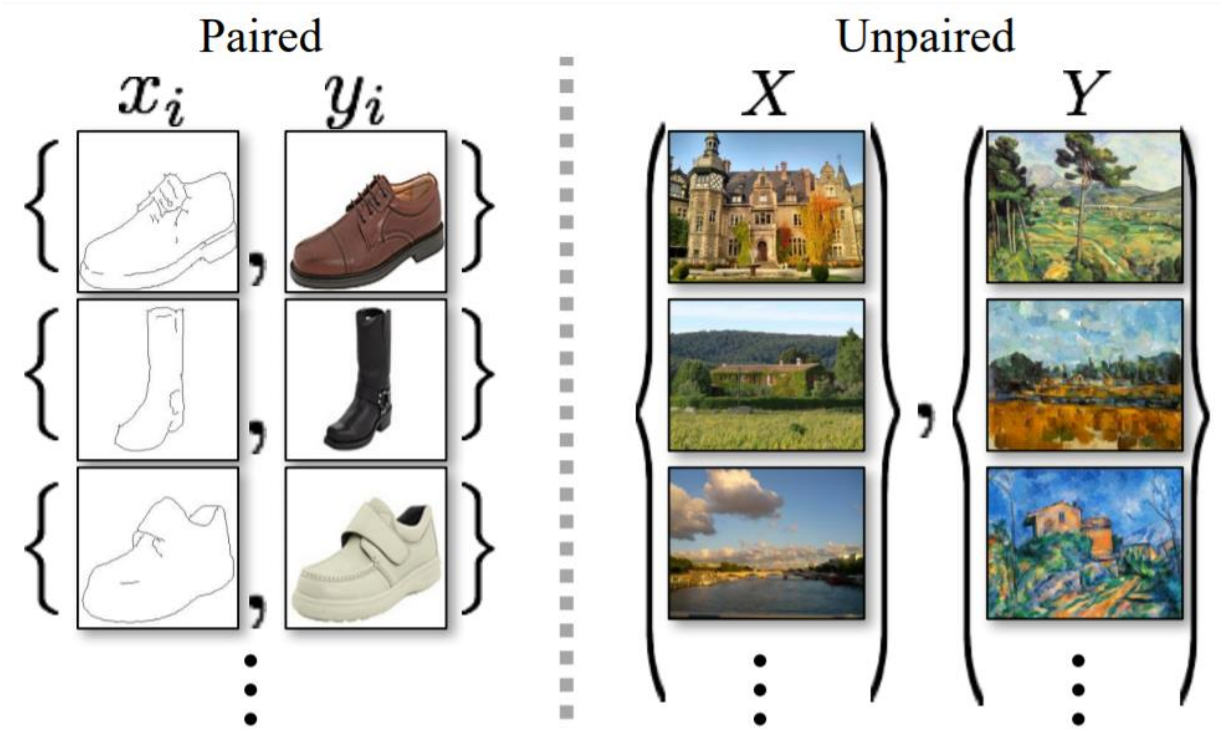
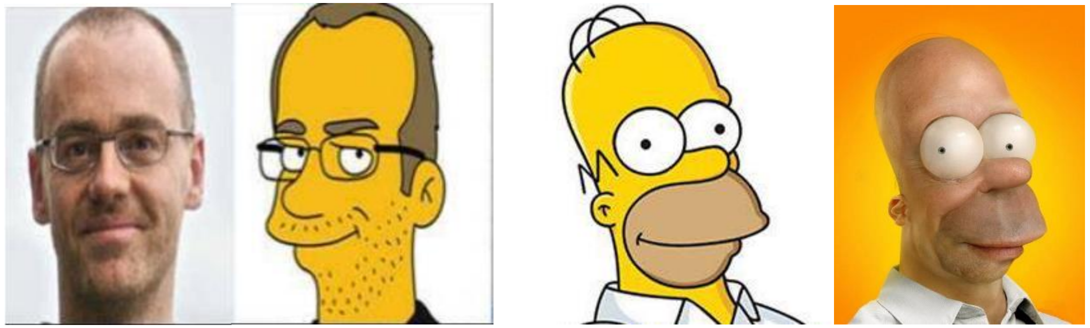
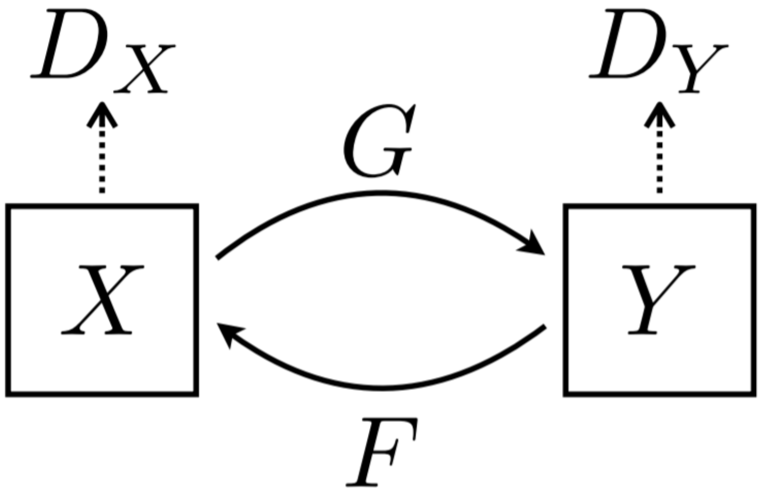


Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- Let's imagine we wanted to build a **Simpsonifyer**
  - Turning photos into "The Simpsons" characters
  - We'd need 2 Generators and 2 Discriminators:
    - **G** – Transform photos (X) into Simpsons (Y)
    - **F** – Transform Simpsons (Y) into photos (X)
    - D_X – Discriminate real vs. fake photos (X)
    - D_Y – Discriminate real vs. fake Simpsons (Y)





Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- **Adversarial Loss** (or GAN loss), for each direction (min-max)
  - Generate realistic images of the target domain
  - $G(X) \rightarrow D_Y$
  - $F(Y) \rightarrow D_X$

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]$$

Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- **Cycle-Consistency Loss**
  - The "magic" of the CycleGAN, ensures better reconstructions (key feature preservation)

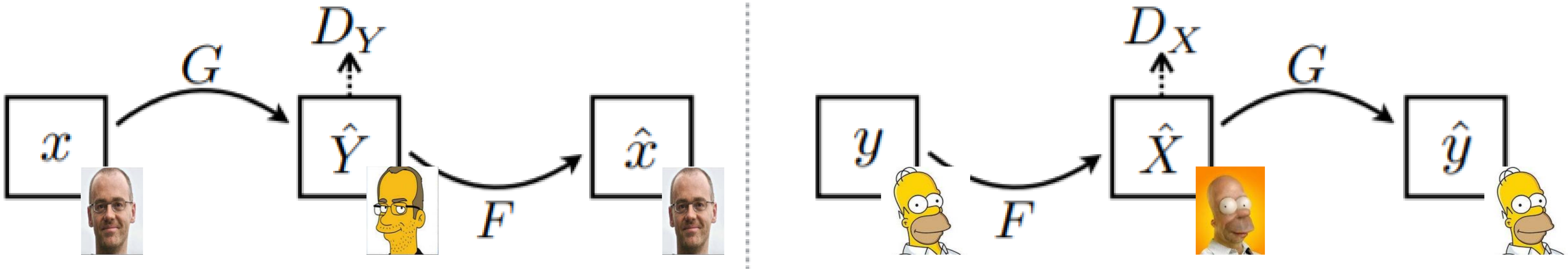
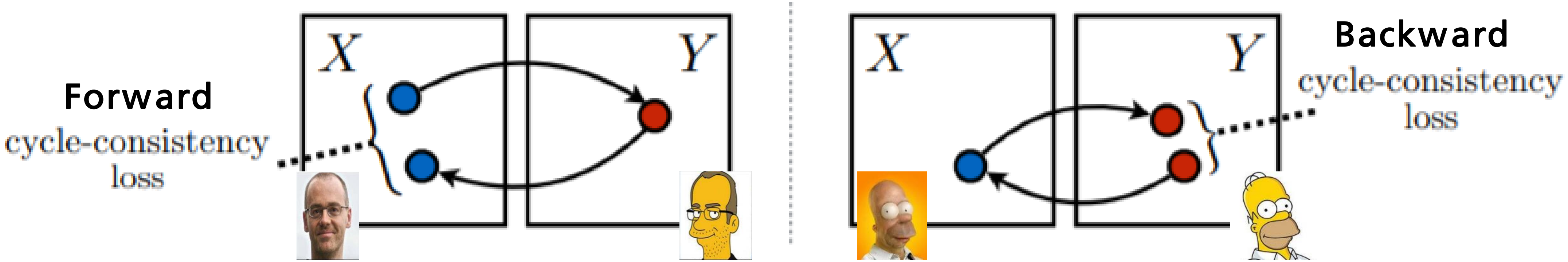


Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- **Cycle-Consistency Loss**
  - The “magic” of the CycleGAN, ensures better reconstructions (key feature preservation)

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$



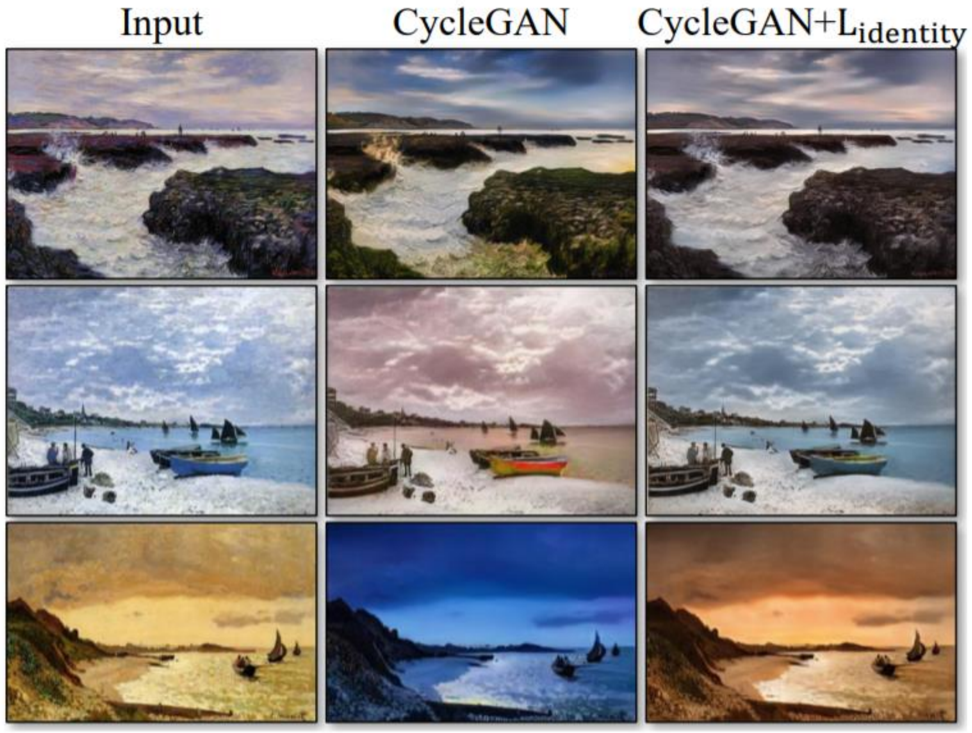
Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- **Identity Loss**
  - It’s an optional loss term, but can help to preserve original colors, textures, etc.
  - Just expects that if a real domain image is passed through the corresponding Generator, it’s still preserved

$$\mathcal{L}_{\text{identity}}(G, F) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(y) - y\|_1] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(x) - x\|_1].$$



Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- Adding all up:

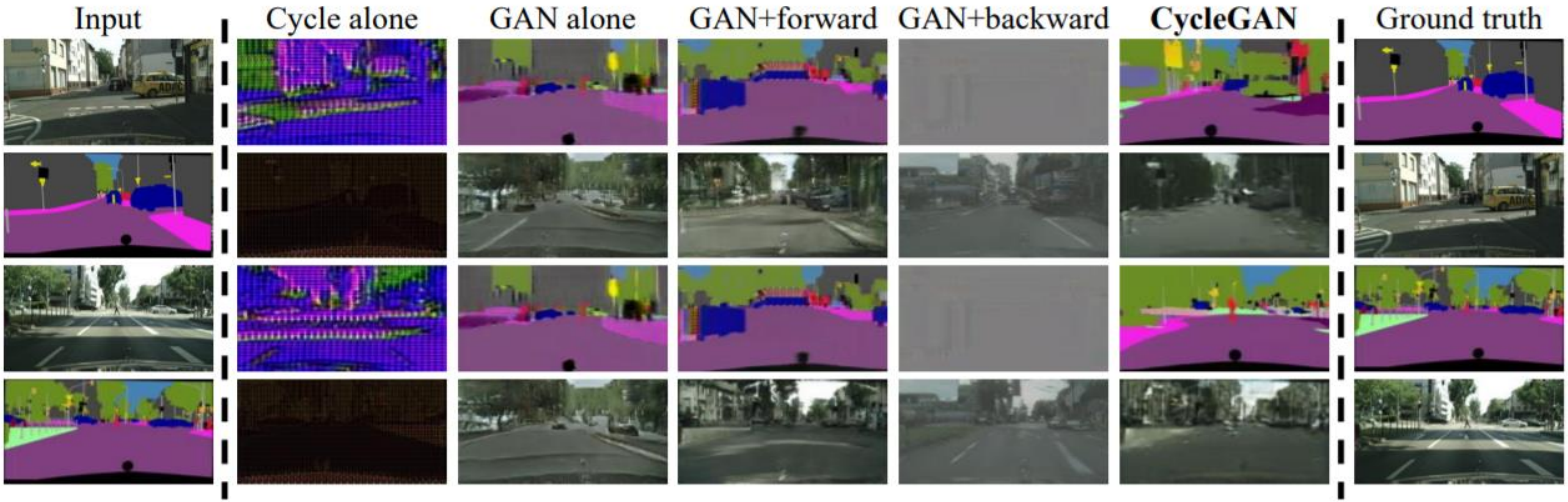
$$
\begin{aligned}
\mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\
& + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\
& + \lambda \mathcal{L}_{\text{cyc}}(G, F), \\
& \boxed{+ \mathcal{L}_{\text{identity}}(G, F)}
\end{aligned}
$$

$$
G^*, F^* = \arg \min_{G,F} \max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y).
$$

Fraunhofer PORTUGAL

Image Translation – CycleGAN

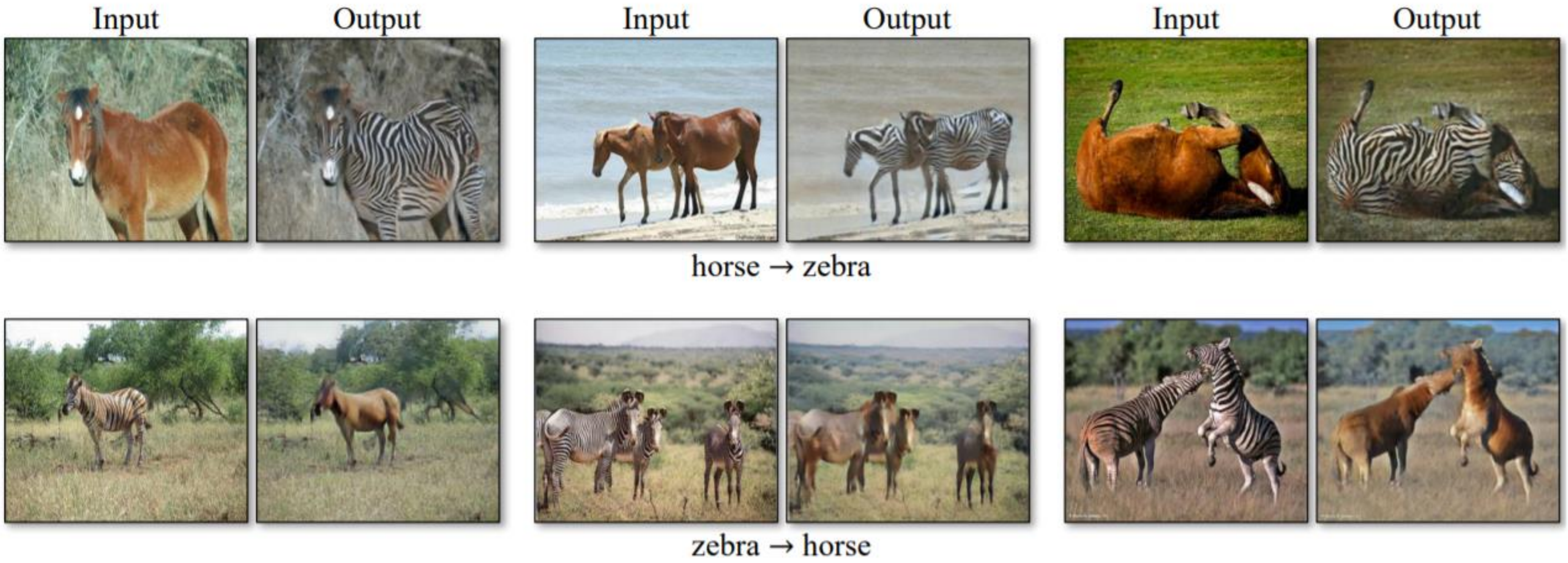
Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.



Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.



Fraunhofer PORTUGAL

Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

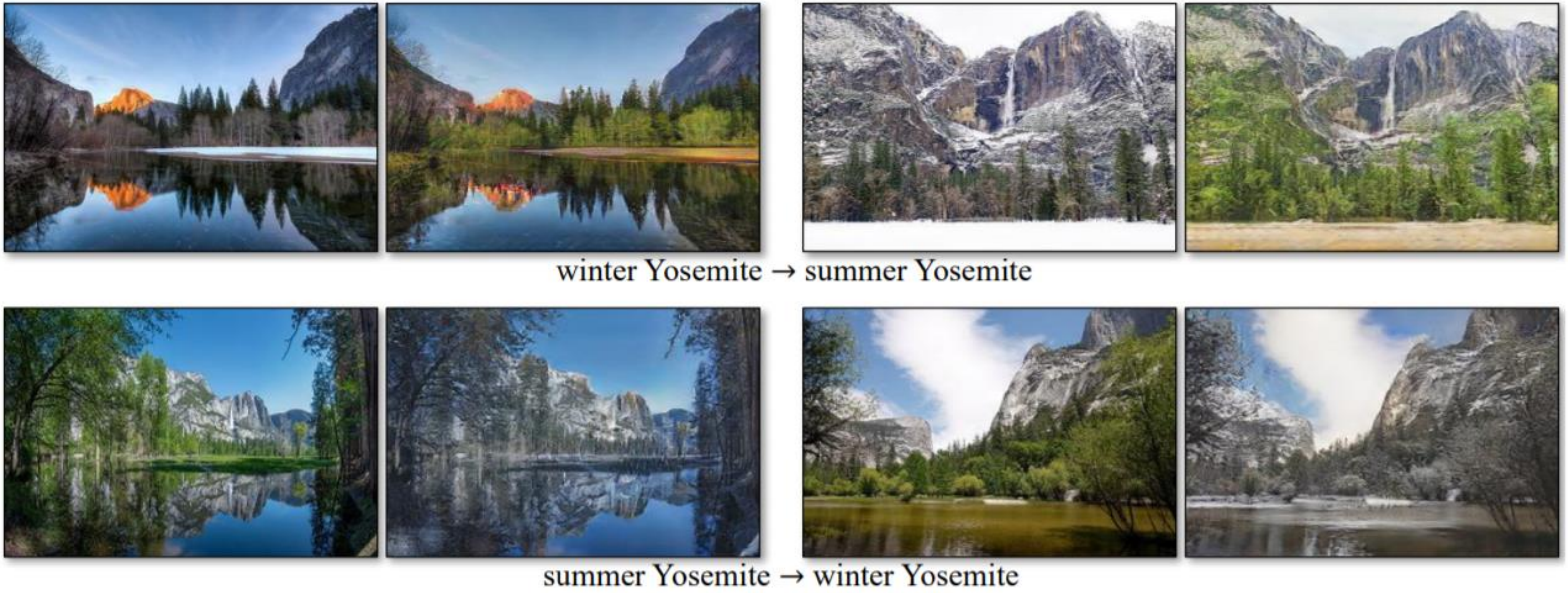
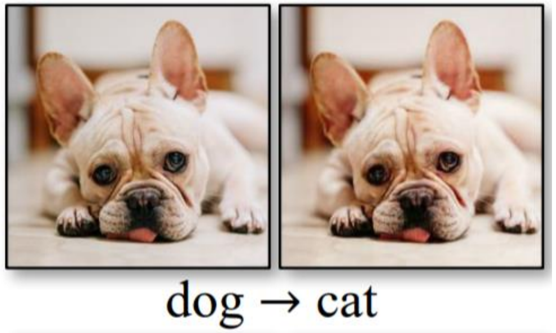
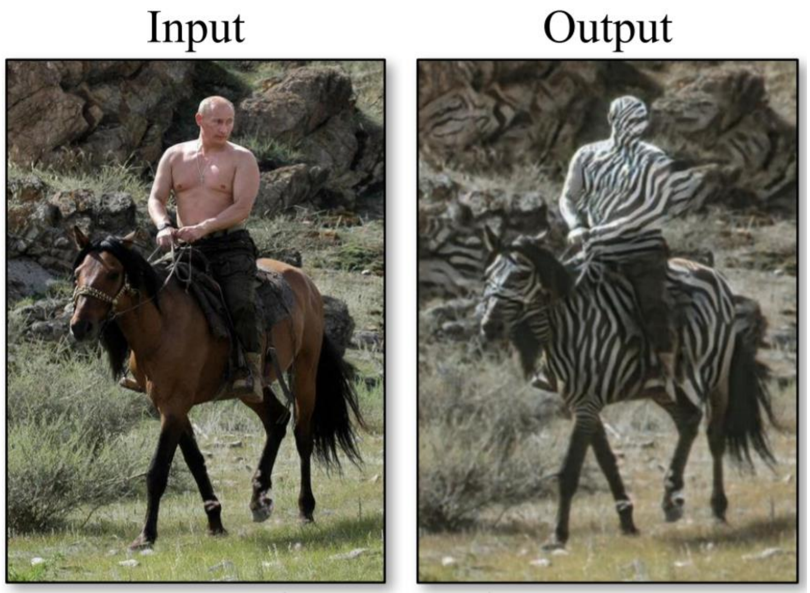


Image Translation – CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks.

- **Limitations**
  - Structural changes are hard, CycleGAN makes minimal input changes
  - CycleGAN does not generalize well to unseen "properties"
    - E.g. no riders in the horses training set





Fraunhofer PORTUGAL

Image Translation – UNIT

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks

- UNIT
  - Instead of relying on explicit cycle-consistency, UNIT (and MUNIT, the multimodal counterpart) propose to learn a shared "content" latent space
    - Relies on
      - GANs for the Generators
      - VAEs for the Encoders (mapping to the latent space)

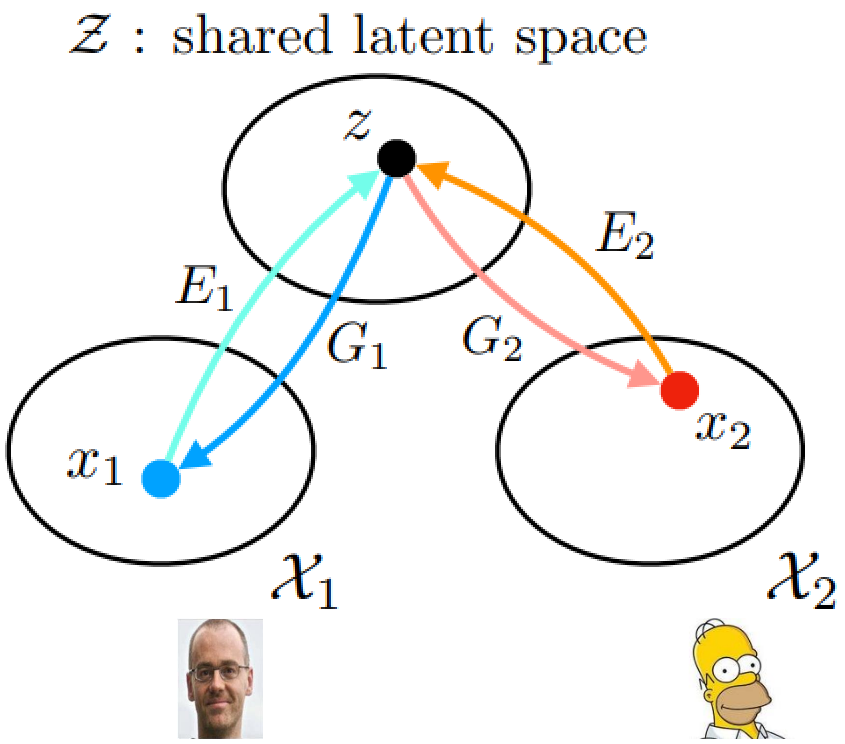


Image Translation – UNIT

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks

The UNIT Framework

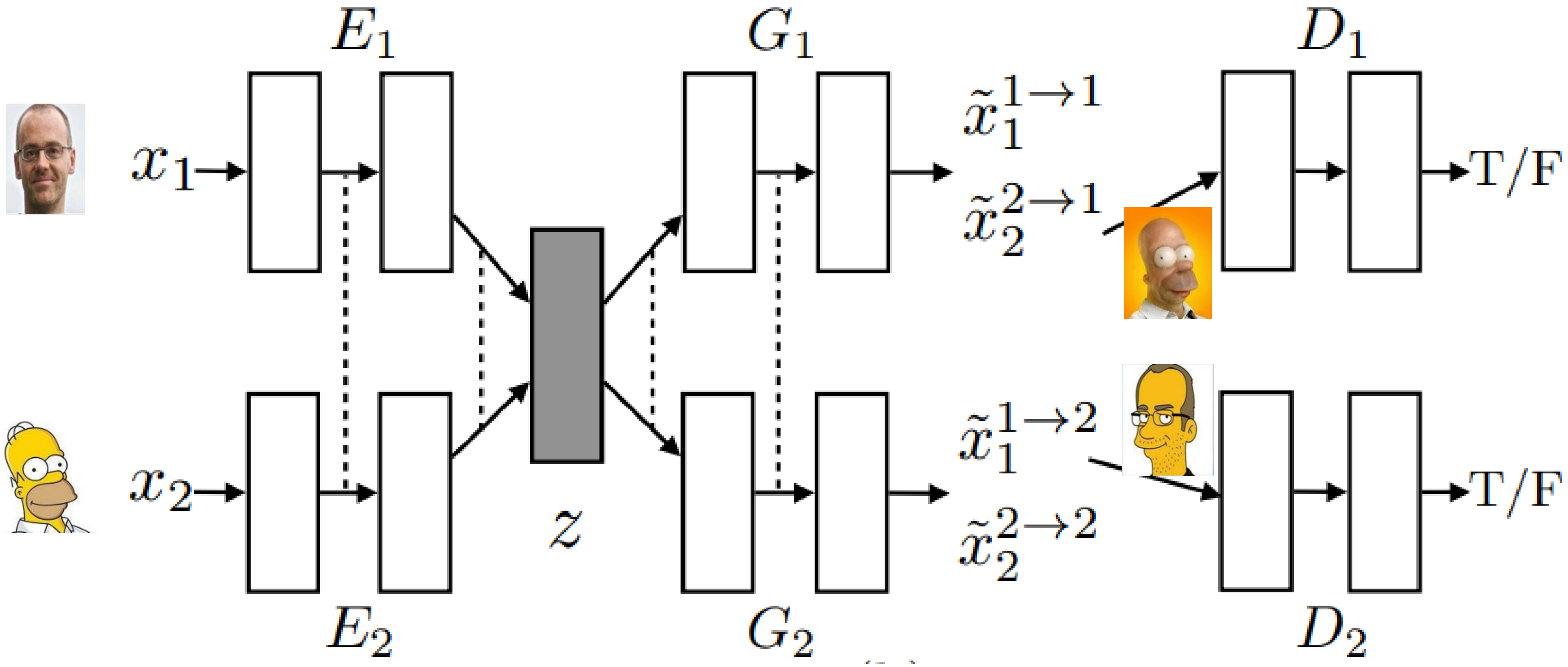


Image Translation – UNIT

Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks

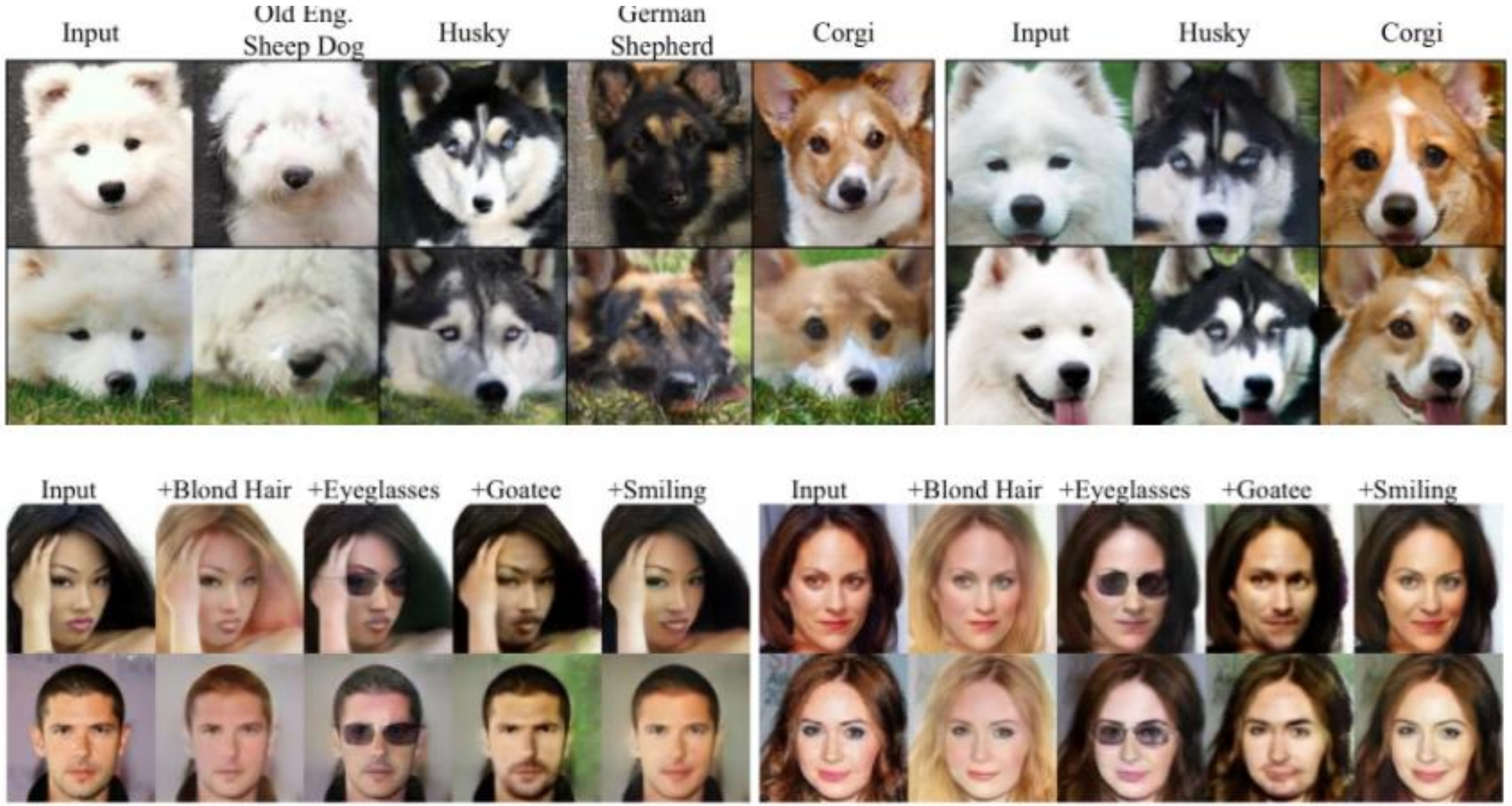


Image Translation – CUT

Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020, August). Contrastive learning for unpaired image-to-image translation.

- **Constrastive Unpaired Translation (CUT)**
  - Constrastive Learning
    - Positives vs. Negatives examples
    - Maximizes Mutual Information between corresponding intput/output patches
  - One-sided Translation
    - Bypasses cycle-consistency
  - Single image translation!
    - Negatives are sampled from same image

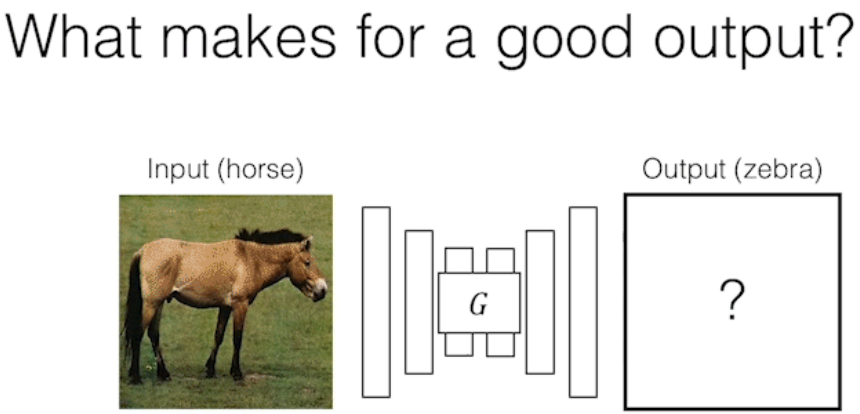


Image Translation – CUT

Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020, August). Contrastive learning for unpaired image-to-image translation.

- **Constrastive Unpaired Translation (CUT)**
  - Reduced training times
  - Generator is broken into two components:
    - G_{ENC} - encoder
    - G_{DEC} - decoder
  - As UNIT, uses a latent space

$$\hat{\boldsymbol{y}} = G(\boldsymbol{z}) = G_{\mathrm{dec}}(G_{\mathrm{enc}}(\boldsymbol{x})).$$

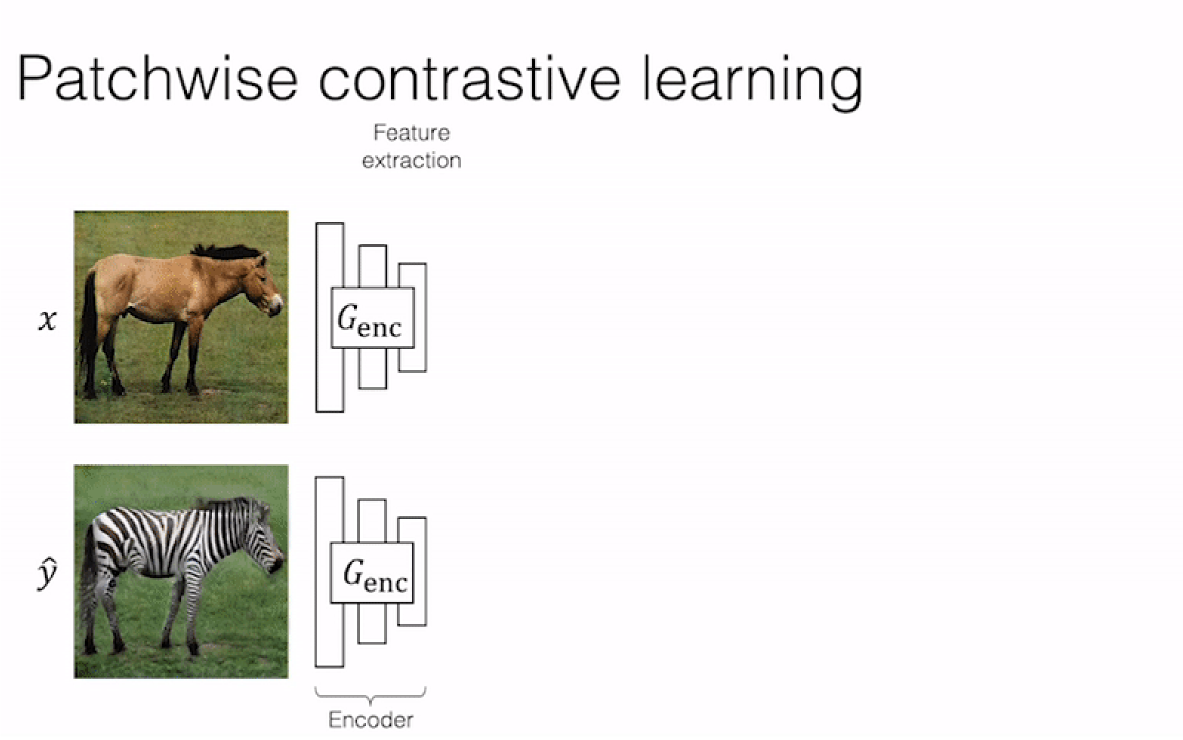


Image Translation – CUT

Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020, August). Contrastive learning for unpaired image-to-image translation.

- Loss Terms
  - Adversarial Loss

$$\mathcal{L}_{\text{GAN}}(G, D, X, Y) = \mathbb{E}_{\boldsymbol{y}\sim Y} \log D(\boldsymbol{y}) + \mathbb{E}_{\boldsymbol{x}\sim X} \log(1 - D(G(\boldsymbol{x})))$$

  - Cross-entropy Loss (probability of the positive example being selected over negatives
  - PatchNCE Loss
    - z_l^s is the feature vector ($H_l(G_{ENC}^l(x))$) for layer l and spatial location s

$$\mathcal{L}_{\text{PatchNCE}}(G, H, X) = \mathbb{E}_{\boldsymbol{x}\sim X} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell(\hat{\boldsymbol{z}}_l^s, \boldsymbol{z}_l^s, \boldsymbol{z}_l^{S\setminus s})$$

  - Adding all up

$$\mathcal{L}_{\text{GAN}}(G, D, X, Y) + \lambda_X \mathcal{L}_{\text{PatchNCE}}(G, H, X) + \lambda_Y \mathcal{L}_{\text{PatchNCE}}(G, H, Y)$$

Image Translation – CUT

Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020, August). Contrastive learning for unpaired image-to-image translation.

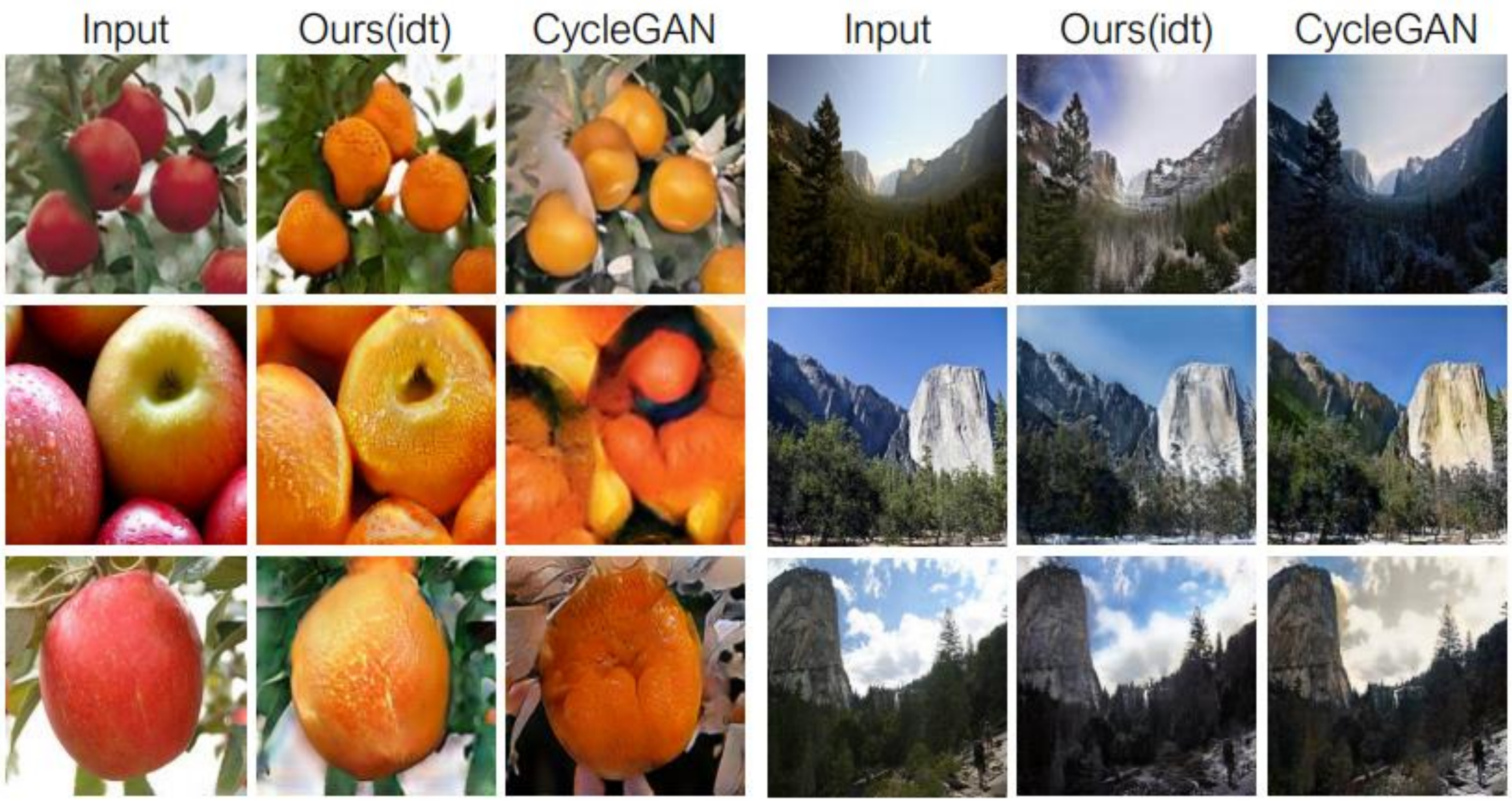
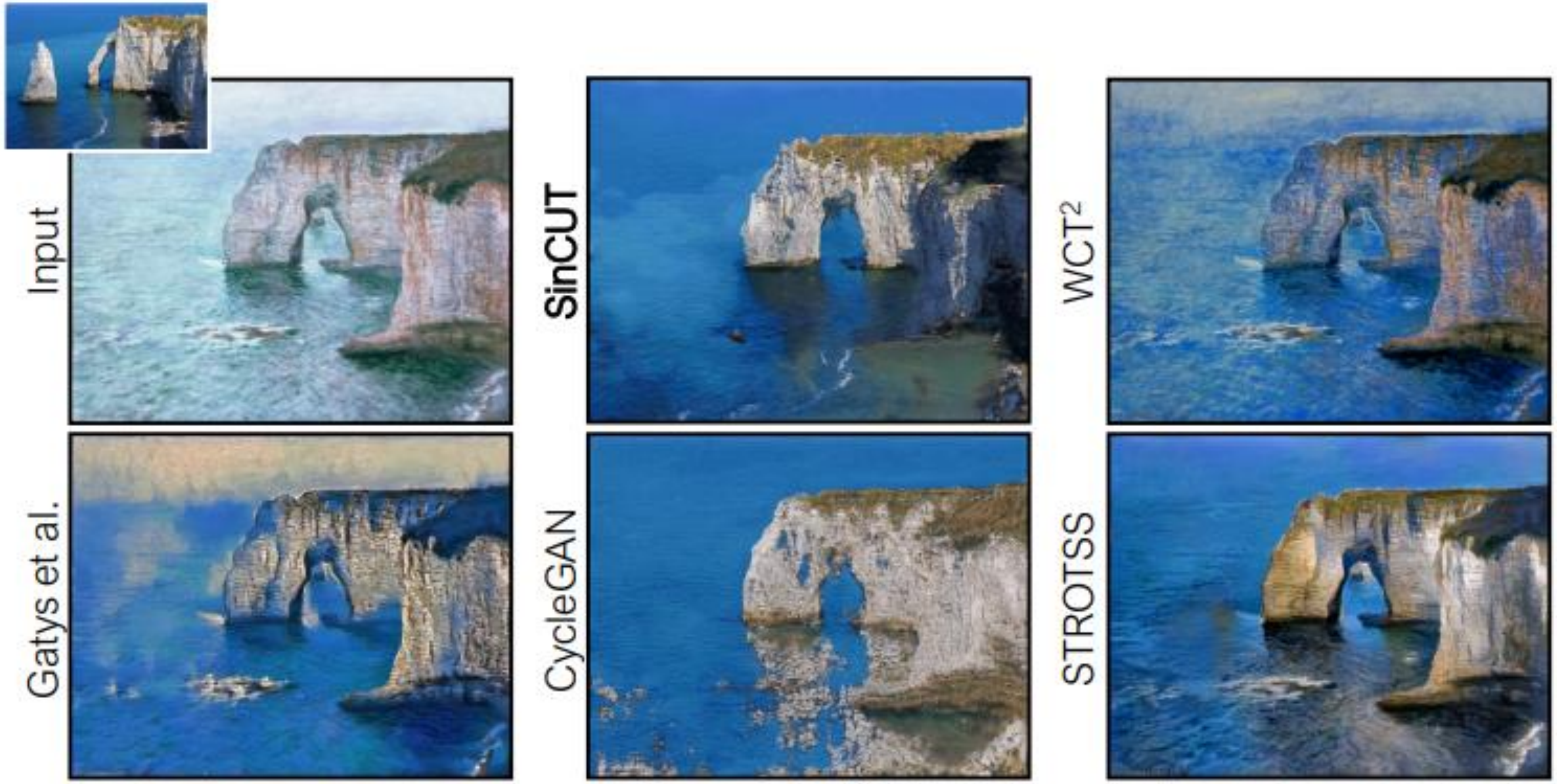


Image Translation – CUT

Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020, August). Contrastive learning for unpaired image-to-image translation.

- Single Image



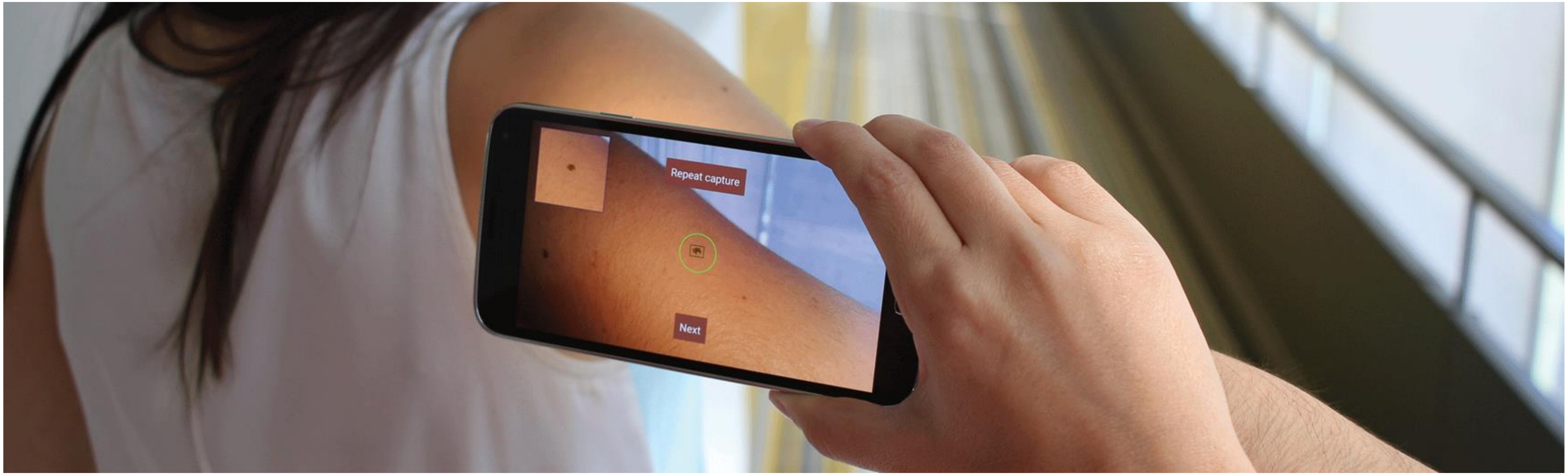
Fraunhofer PORTUGAL

GANS at Work @ Fraunhofer Portugal

Not just Cats, Zebras and Simpsons

Improving Dermatological Diagnosis

Catarina Andrade et al. (2020)



Improving Dermatological Diagnosis

Catarina Andrade et al. (2020)

- The main goal is to render Teledermatology a reality
  - Segment skin lesions using a smartphone camera - macroscopic images)
  - However, there are little annotated datasets in this domain
  - Several datasets with dermoscopic images available

Acquisition device:

Dermoscope

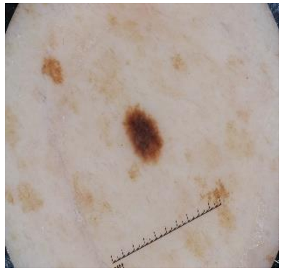
Image Characteristics:

Structures and features present

Colours and pigmentation

Artefacts

Flat outward aspect



≠

Acquisition device:

Digital Camera or Smartphone

Image Characteristics:

Surface Glare

Reflections

Visual Depth

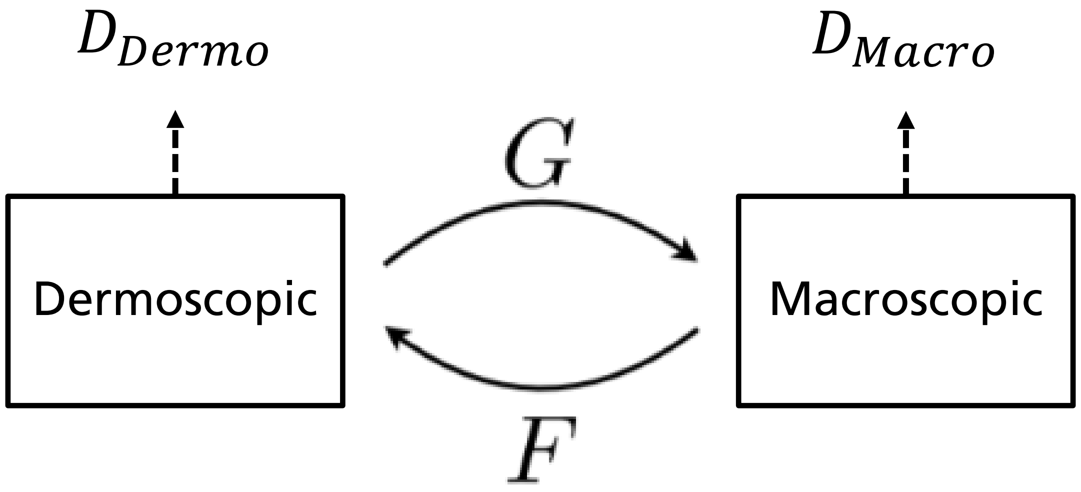
Decreased Resolution

Fraunhofer PORTUGAL

Improving Dermatological Diagnosis

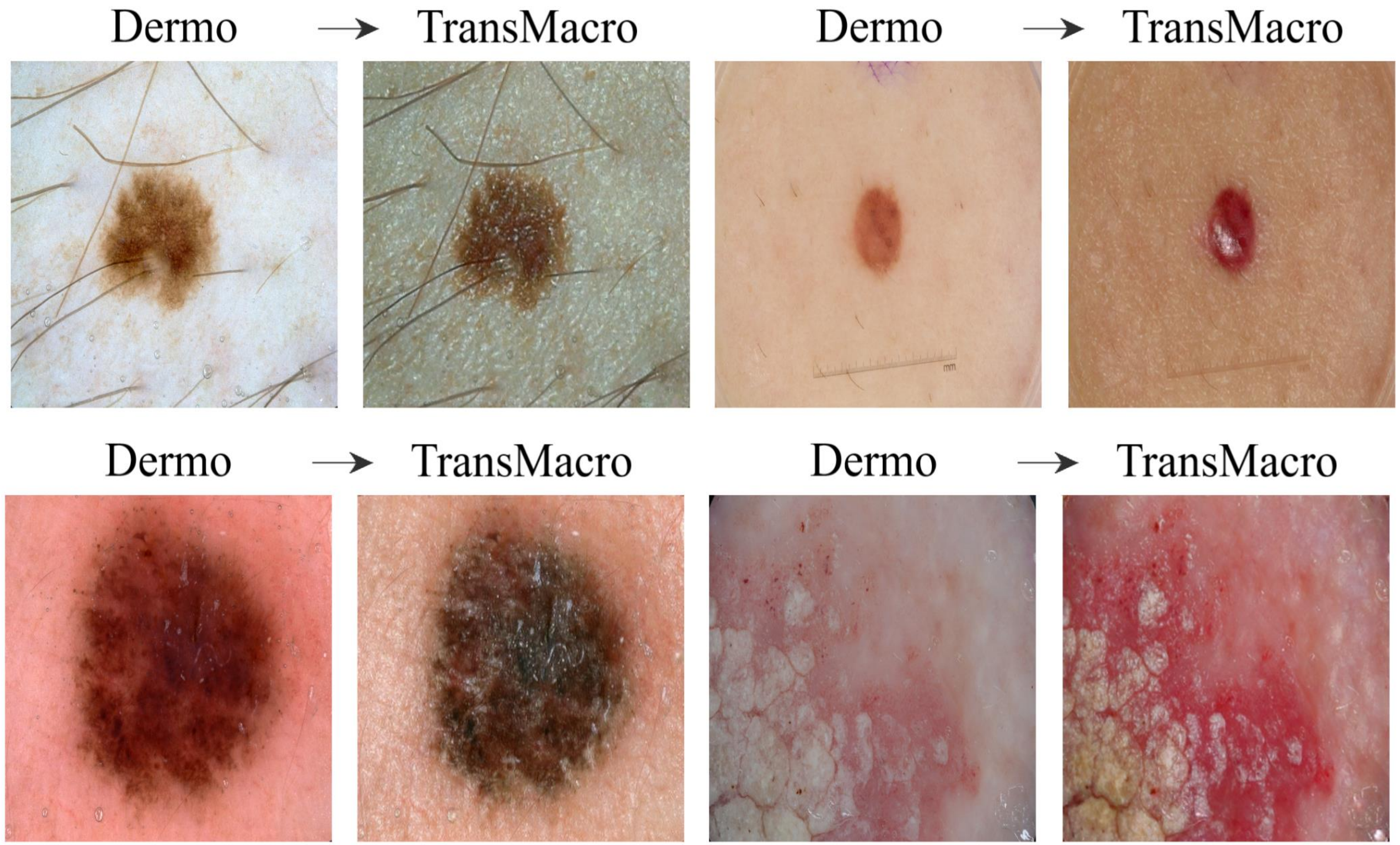
Catarina Andrade et al. (2020)

- Idea:
  - Learn to segment lesions using the Dermoscopic Datasets
    - U-Net and DeepLab with MobilenetV2 encoder (lightweight)
  - Use CycleGAN to map from Dermoscopic to Macroscopic domain and enrich training



Improving Dermatological Diagnosis

Catarina Andrade et al. (2020)



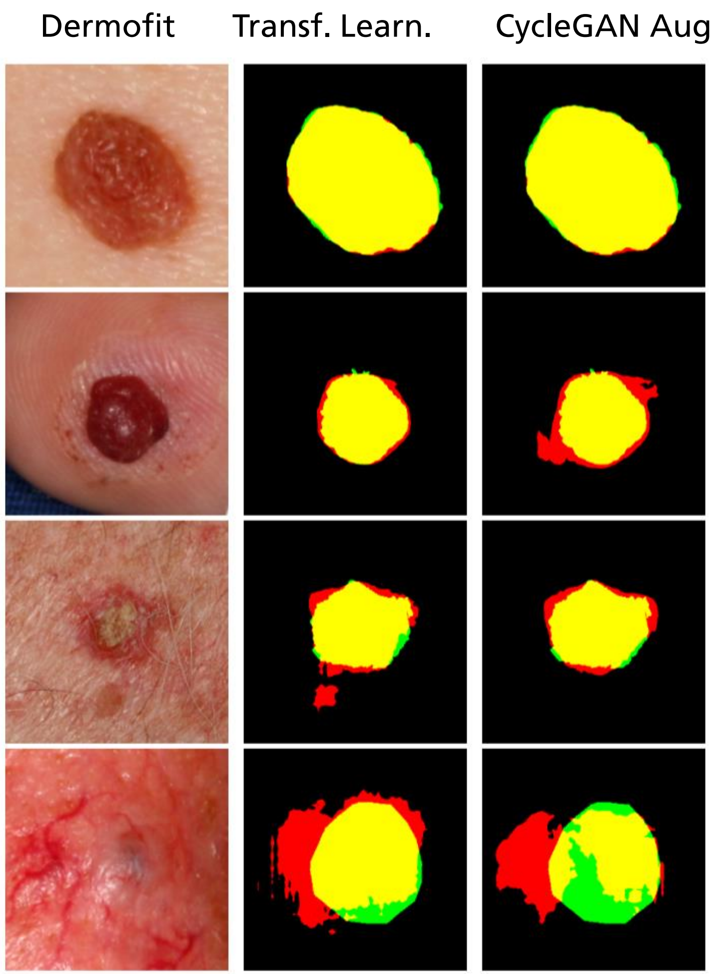
Improving Dermatological Diagnosis

Catarina Andrade et al. (2020)

Dermofit

Experiment	Thres	TJA	JA	DI	AC	SE	SP
Transfer Learning	0.50	72.97	80.26	88.26	93.51	87.56	**96.86**
CycleGAN Augment.	**0.56**	**75.46**	**81.03**	**88.79**	**93.78**	**89.68**	96.13

Outperforms Transf. Learn. by 2.49% TJA and 0.77% JA



Yellow – True Positives; Red – False Positive; Green – False Negative

Improving the Quality of Retinal Images

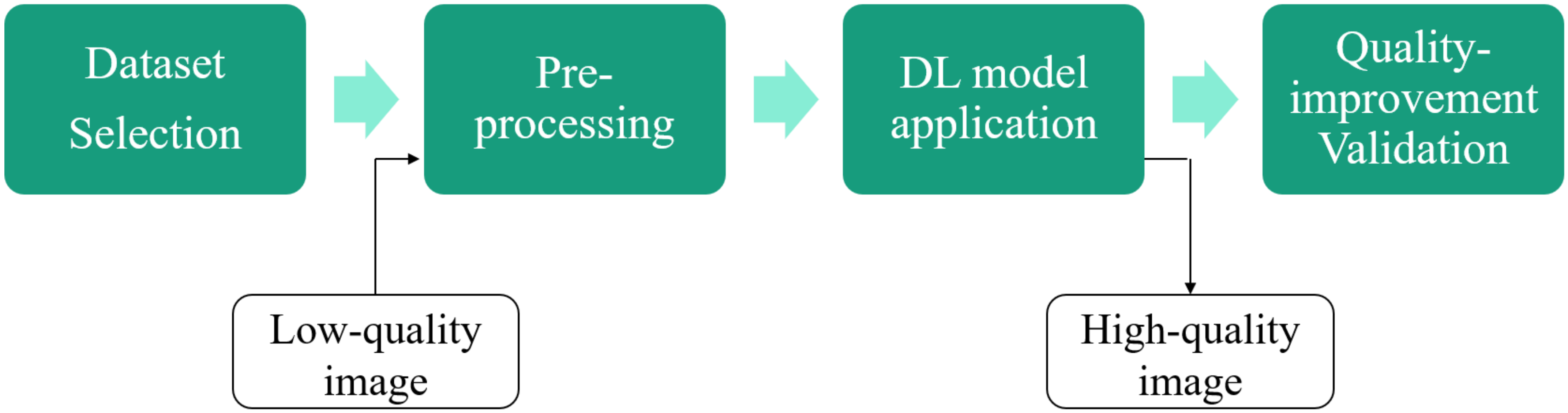
- Meet the **EyeFundusScope**
  - Affordable
  - Portable
  - Non-Experts can use

  - Still lags behind in image quality when compared to much more expensive and heavy table-top equipment

Improving the Quality of Retinal Images

Beatriz Simões et al. (2020)

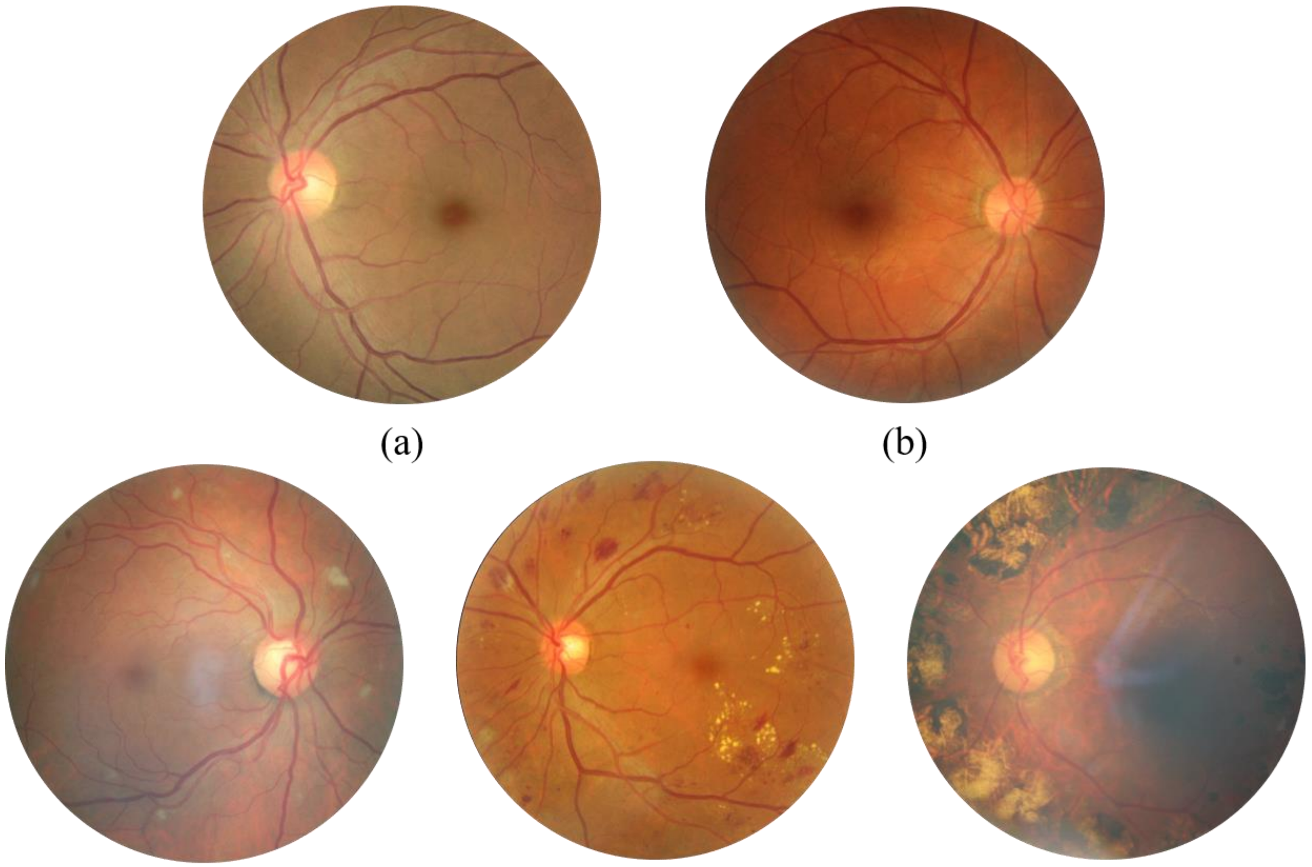
- The Pipeline



Improving the Quality of Retinal Images

The Data

- Datasets
  - Eyepacs – Diabetic Retinopathy (DR)
    - > 5M retina images



(a) (b)

(c) (d) (d)

(a) no DR;
(b) mild DR;
(c) moderate DR;
(d) severe DR;
(e) proliferative DR

Fraunhofer PORTUGAL

Improving the Quality of Retinal Images

The Data

- Datasets
  - EyeQual – Image Quality Annotations

(a) (b) (c)

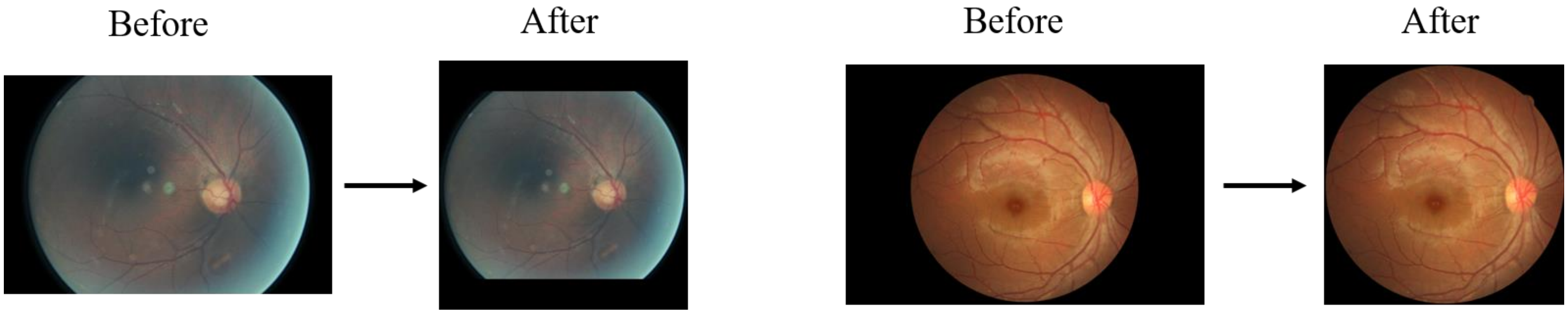
(a) good quality; (b) usable quality; and (c) rejectable quality

Fraunhofer PORTUGAL

Improving the Quality of Retinal Images

Preprocessing

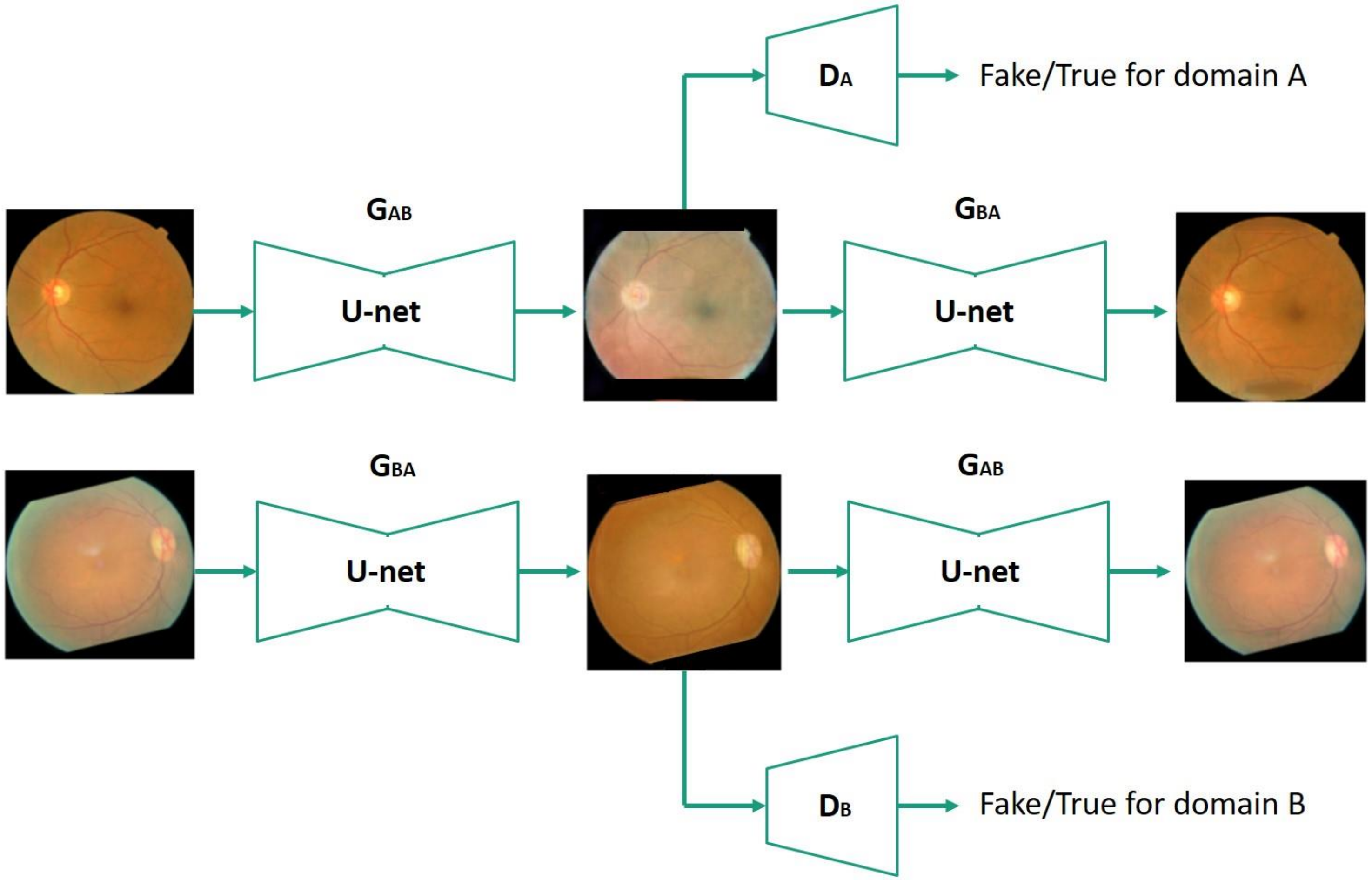
- Uniform Aspect Ratios
- Normalization
- Resizing



Improving the Quality of Retinal Images

Model

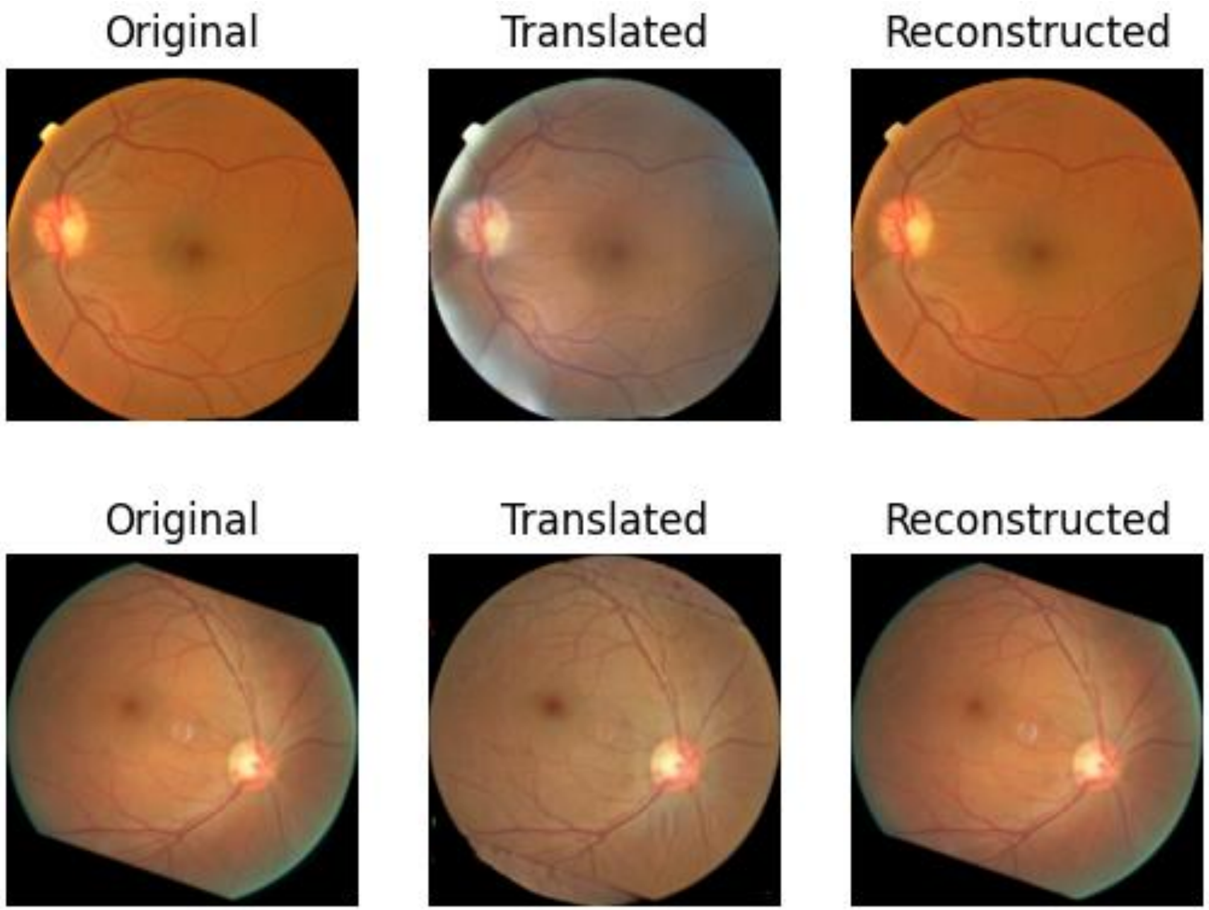
- CycleGAN



Improving the Quality of Retinal Images

Evaluation

- CycleGAN
- The obvious, and still the best, way to evaluate GAN performance is by human judgment
  - Unfortunately we couldn't get expert validation for now, but we have plans for that

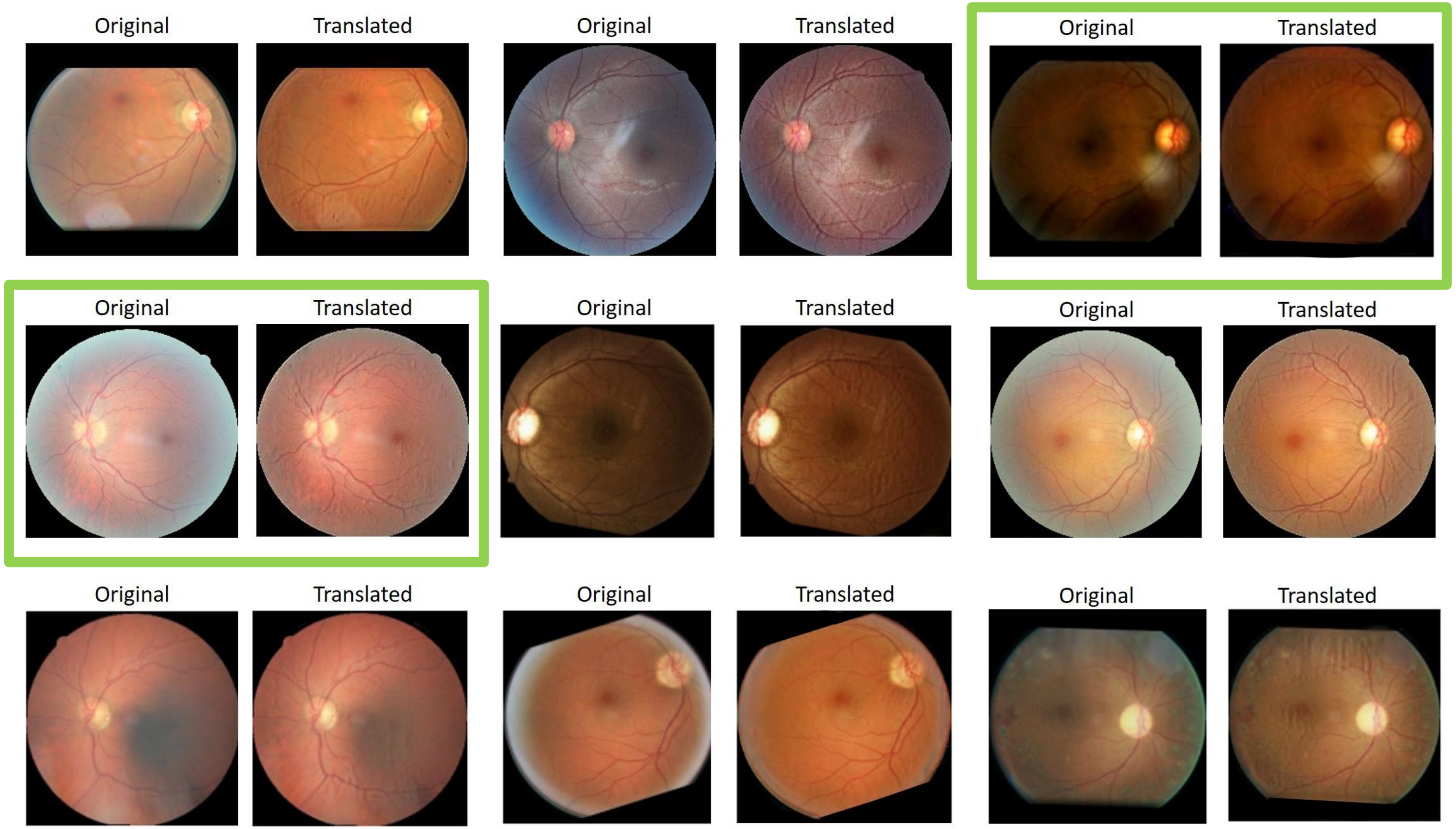


Fraunhofer PORTUGAL

Improving the Quality of Retinal Images

Evaluation

- CycleGAN
  - Agnostic to defects

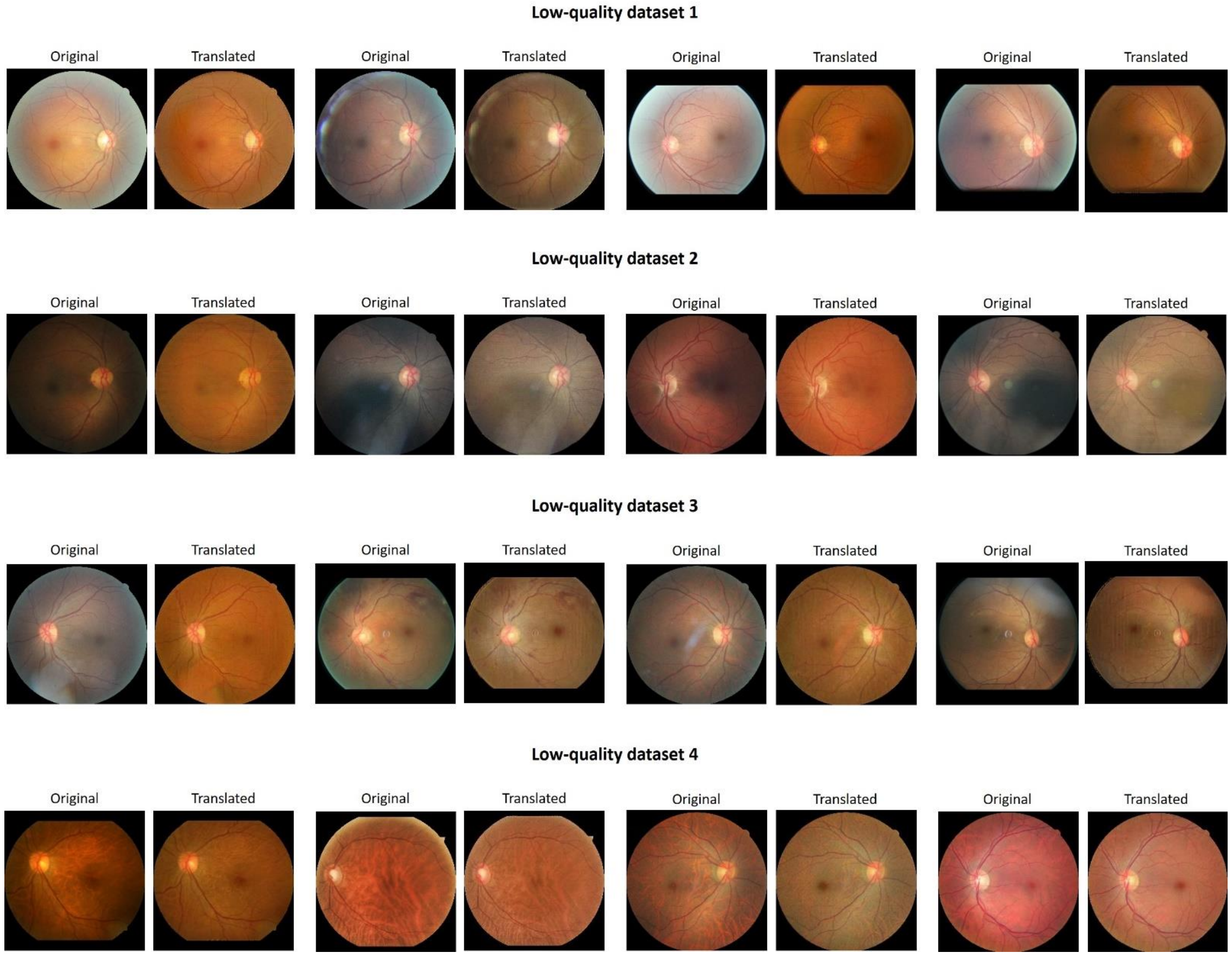


Fraunhofer PORTUGAL

Improving the Quality of Retinal Images

Evaluation

- **CycleGAN**
  - Defect-specific

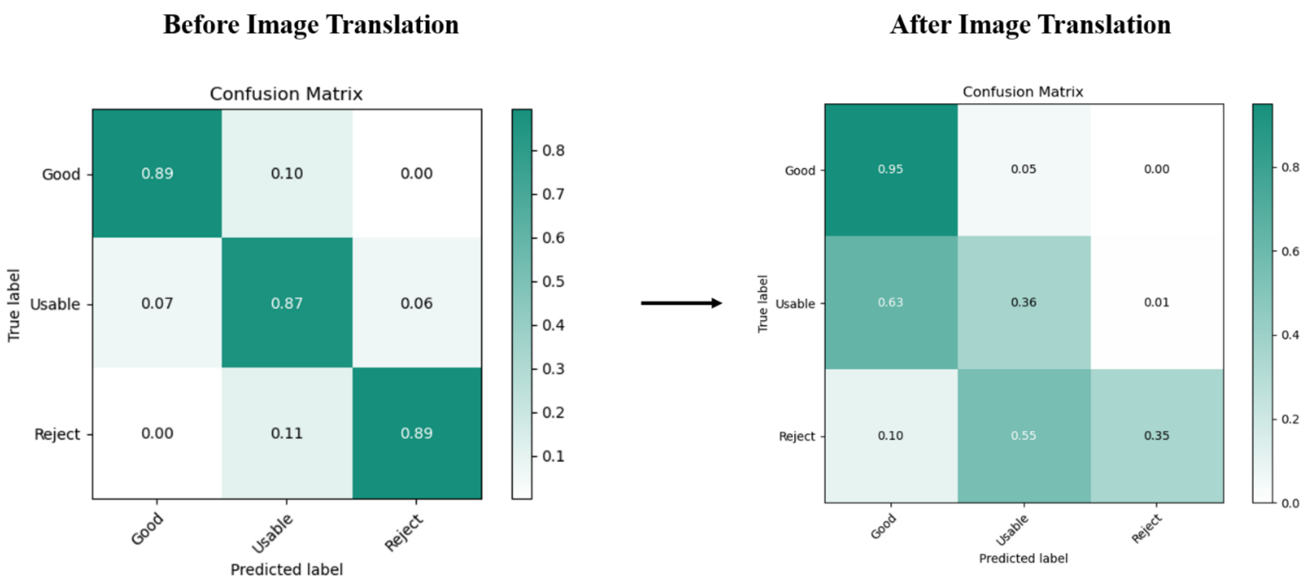


Improving the Quality of Retinal Images

Evaluation

- Standard evaluation methods such as Inception Score (IS) or Frechet Inception Distance (FID) were not appropriate (objects dramatically different from ImageNet pre-training)

- Solution was to use an Image Quality Evaluator
  - Trained with EyeQual Dataset
    - F1-score ~86%
  - Assess the shift towards better Quality levels according to the Evaluator



Fraunhofer PORTUGAL

Improving the Quality of Retinal Images

Evaluation

- Other way to evaluate the impact is to assess how the transformation affects Diagnostic performance of a DR (and recently a Glaucoma) classifier
  - Based on EfficientNet B0 – as the name suggests, is very efficient yet with good performances
  - DR
    - Similar results when "filtering" the images: ~ 73% Cohen's Kappa coefficient
    - However, we found a correlation between image quality and DR severity that our transformer decreases!
  - Glaucoma
    - Results yet to publish but using both improved and degraded retinal images improves the classifiers' performance!

Fraunhofer PORTUGAL

André V. Carreiro

andre.carreiro@fraunhofer.pt

PORTO – Headquarters

Address: **Rua Alfredo Allen 455/461**

4200-135 Porto | Portugal

Phone: **+351 220 430 300**

LISBOA – Branch Office

Address: **Avenida Prof. Gama Pinto 2**

1649-003 Lisbon | Portugal

Website: **www.fraunhofer.pt** | E-mail: **info@fraunhofer.pt** | Facebook: **facebook.com/fraunhoferportugal** | LinkedIn: **Fraunhofer Portugal**

Fraunhofer PORTUGAL