Auxiliary tasks for the conditioning of Generative Adversarial Networks

Cyprien Ruffino[1]

April 20, 2021

[1]Normandie Univ, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, 76 000 Rouen, France

Context

Photo-realistic image generation works


http://www.whichfaceisreal.com/

Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." CVPR2019

Photo-realistic image generation works

(the real one is the left one)

http://www.whichfaceisreal.com/

Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." CVPR2019

Applications: Filters

Snapchat / Instagram filters

Rameen Abdal, et al. "Image2StyleGAN++: How to Edit the Embedded Images?", CVPR2020

Applications: DeepFakes

Deepfakes, video falsification

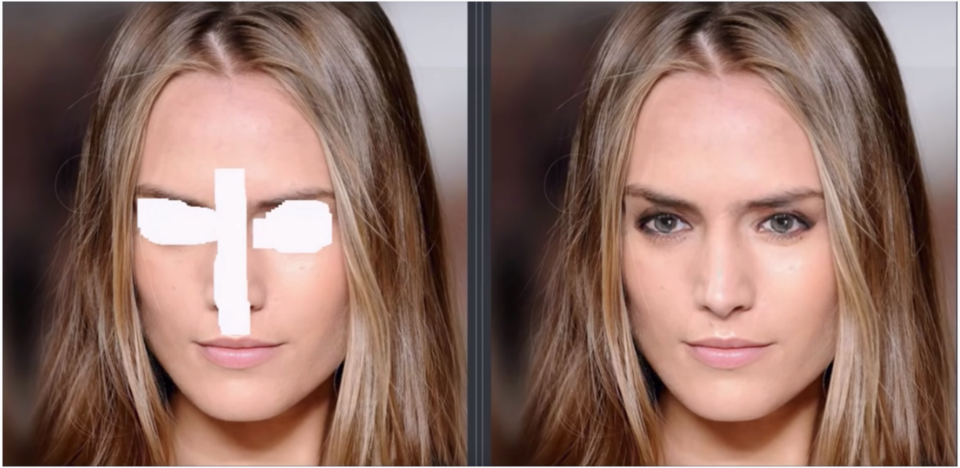
https://www.youtube.com/watch?v=j_LuZlg6xXU

Applications: Image colorization

Image restoration, coloration

Xuan Luo, et al. "Time-Travel Rephotography", on Arxiv, 2020

Applications: Inpainting



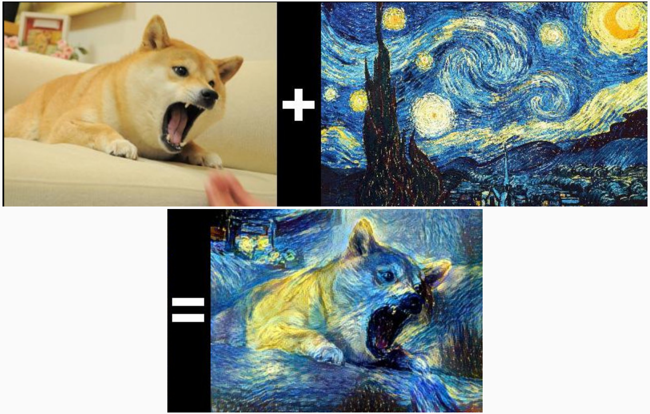
Live photo edition

Liu Guilin, et al. "Image inpainting for irregular holes using partial convolutions." ECCV2018

Applications: Style transfer

Yijun Li, et al. "A Closed-form Solution to Photorealistic Image Stylization", ICCV2017

Applications: Image transfiguration



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

Generative Adversarial Networks

Context

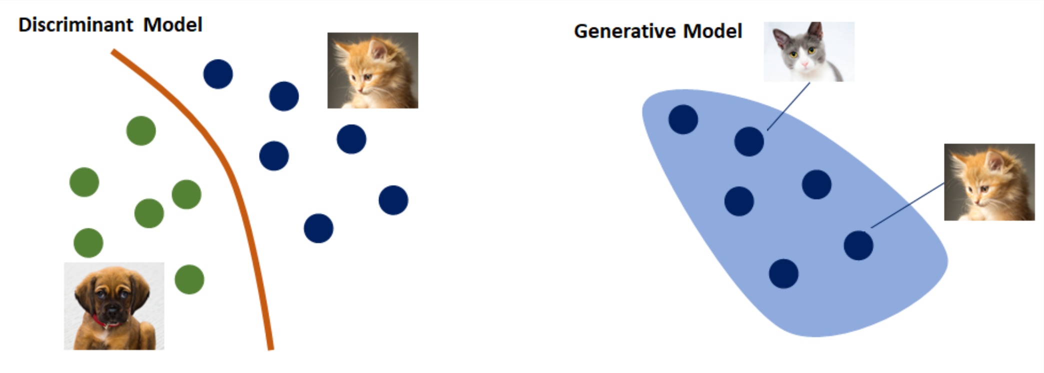
Generative Adversarial Networks

Image reconstruction

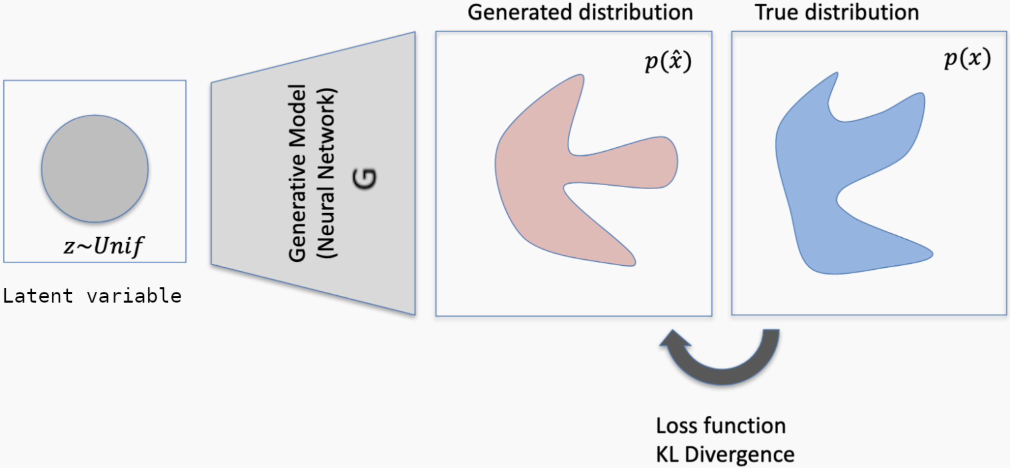
Data augmentation of polarimetric datasets

Conclusion

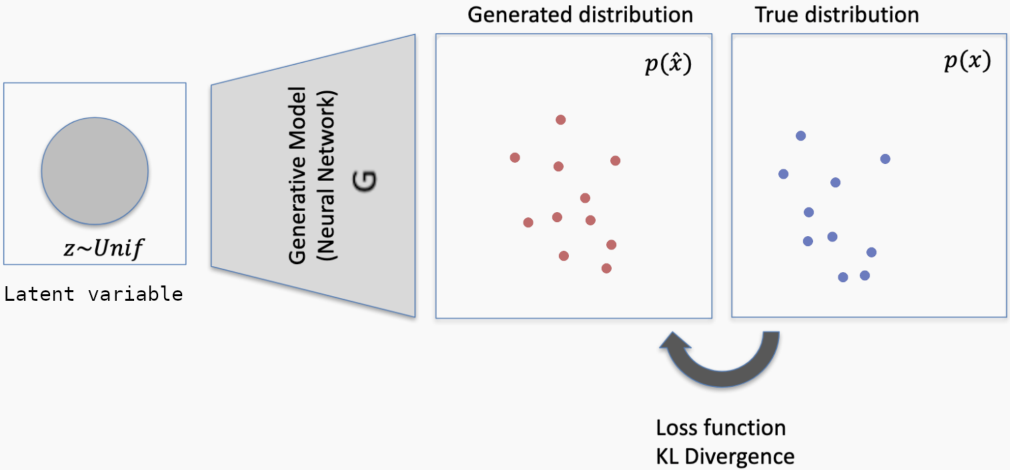
Generative modeling



Latent-variable models



Latent-variable models

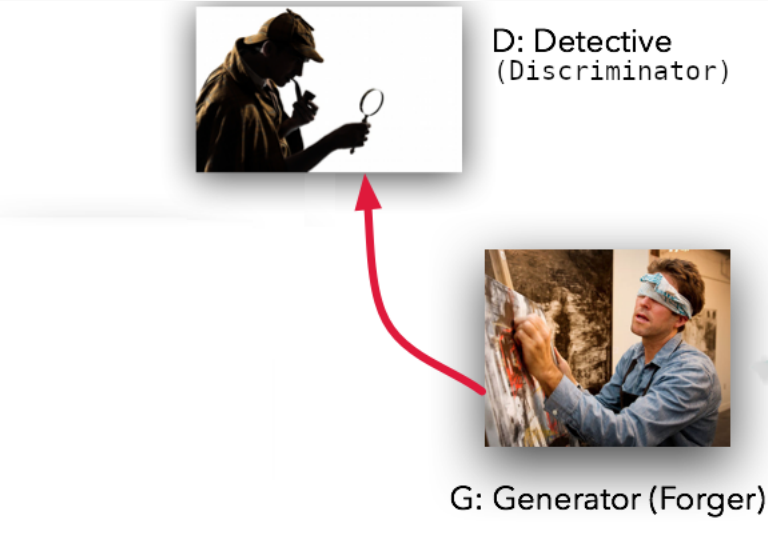


Problem: We can't access the distributions

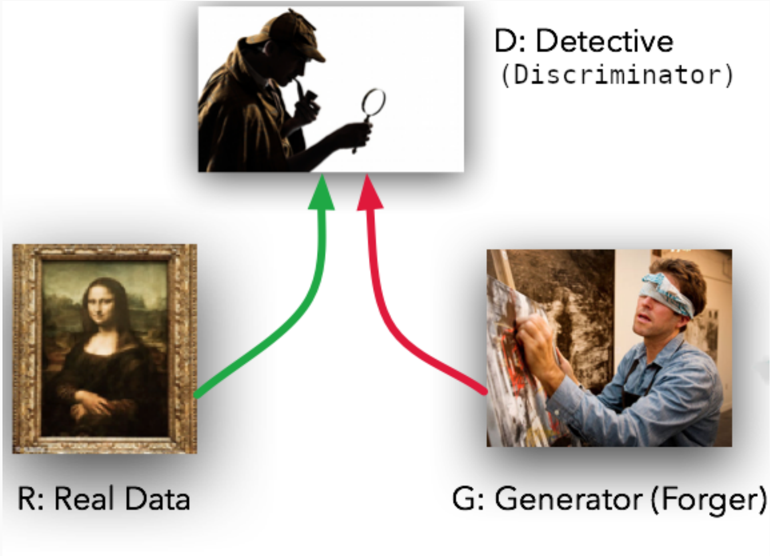
An analogy

G: Generator (Forger)

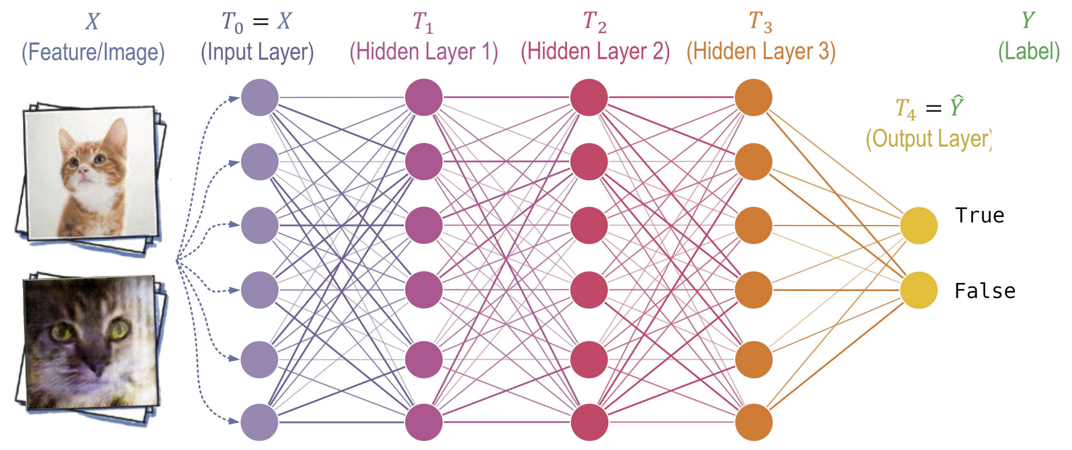
An analogy



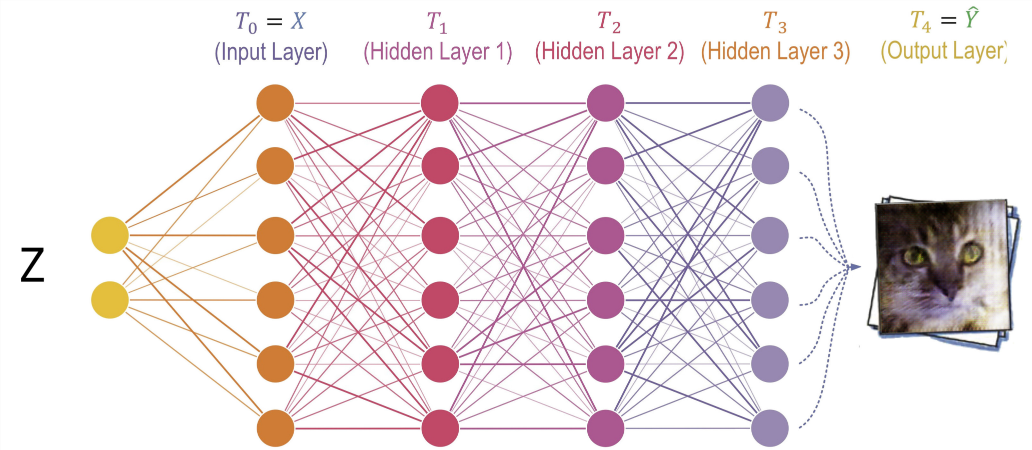
An analogy



Discriminator



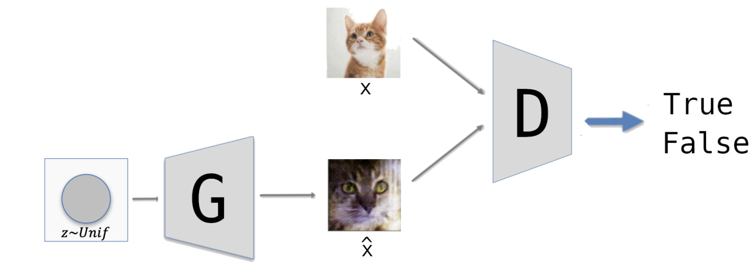
Generator


Z
$T_0 = X$
(Input Layer)
T_1
(Hidden Layer 1)
T_2
(Hidden Layer 2)
T_3
(Hidden Layer 3)
$T_4 = \hat{Y}$
(Output Layer)

Generative Adversarial Networks

Two networks, a generator G and a discriminator D :

- **Generator**: produces synthetic data from a random $z \sim p_Z$, where p_Z is a known distribution
- **Discriminator**: binary classifier, tries to distinguish real samples from fake ones



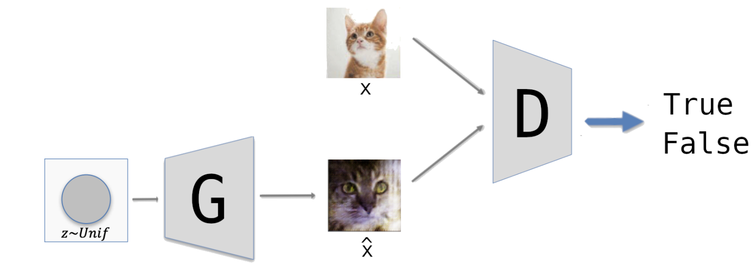
Ian Goodfellow, et al. "Generative adversarial nets." NeurIPS2014

Generative Adversarial Networks

Two networks, a generator G and a discriminator D :

- **Generator**: produces synthetic data from a random $\mathbf{z} \sim p_Z$, where p_Z is a known distribution
- **Discriminator**: binary classifier, tries to distinguish real samples from fake ones

$$\min_{G} \max_{D} L(D, G) = \mathop{\mathbb{E}}_{\mathbf{x} \sim p_X} \left[\log(D(\mathbf{x})) \right] + \mathop{\mathbb{E}}_{\mathbf{z} \sim p_Z} \left[\log(1 - D(G(\mathbf{z}))) \right]$$

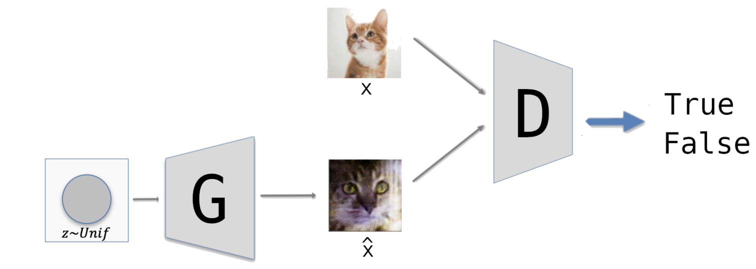


Generative Adversarial Networks

Two networks, a generator G and a discriminator D :

- **Generator**: produces synthetic data from a random $\mathbf{z} \sim p_Z$, where p_Z is a known distribution
- **Discriminator**: binary classifier, tries to distinguish real samples from fake ones

$$\min_G \max_D L(D, G) = \mathop{\mathbb{E}}_{\mathbf{x} \sim p_X} \Big[\log(D(\mathbf{x})) \Big] + \mathop{\mathbb{E}}_{\mathbf{z} \sim p_Z} \Big[\log(1 - D(G(\mathbf{z}))) \Big]$$

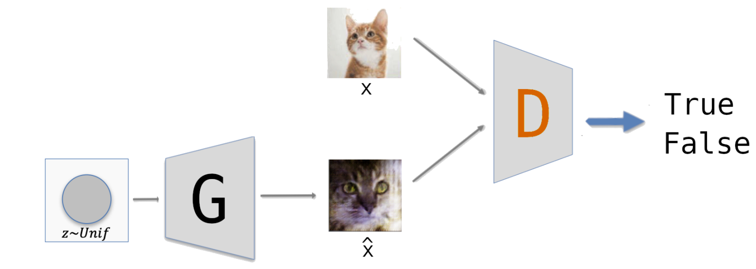


Generative Adversarial Networks

Two networks, a generator G and a discriminator D :

- **Generator**: produces synthetic data from a random $\mathbf{z} \sim p_Z$, where p_Z is a known distribution
- **Discriminator**: binary classifier, tries to distinguish real samples from fake ones

$$\min_G \max_D L(D, G) = \mathop{\mathbb{E}}_{\mathbf{x} \sim p_X} \left[\log(D(\mathbf{x})) \right] + \mathop{\mathbb{E}}_{\mathbf{z} \sim p_Z} \left[\log(1 - D(G(\mathbf{z}))) \right]$$

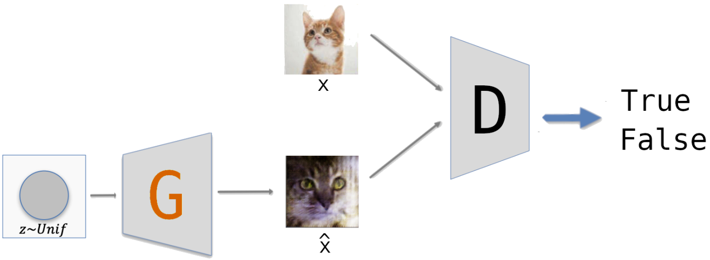


Generative Adversarial Networks

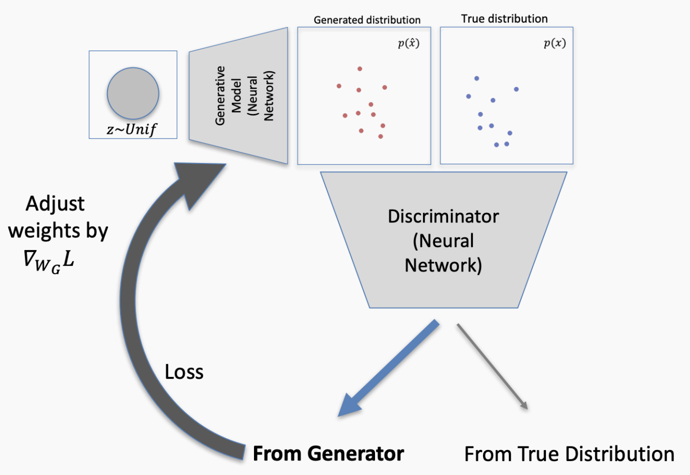
Two networks, a generator G and a discriminator D :

- **Generator**: produces synthetic data from a random $\mathbf{z} \sim p_Z$, where p_Z is a known distribution
- **Discriminator**: binary classifier, tries to distinguish real samples from fake ones

$$\min_G \max_D L(D, G) = \mathbb{E}_{\mathbf{x} \sim p_X} \left[\log(D(\mathbf{x})) \right] + \mathbb{E}_{\mathbf{z} \sim p_Z} \left[\log(1 - D(G(\mathbf{z}))) \right]$$

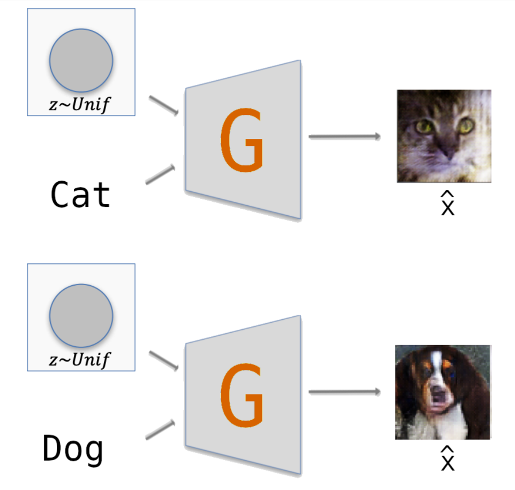


GAN training



Controlling the generation

Controlling the generation



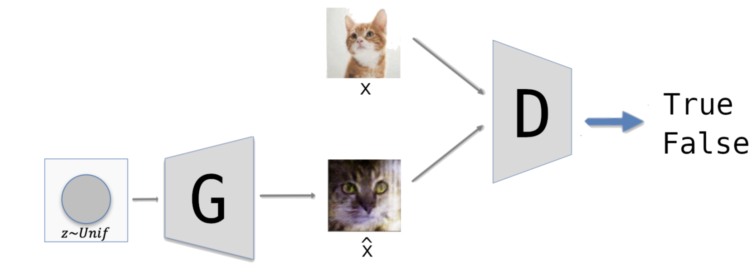
Adding labels to the GAN

Generative Adversarial Networks

Two networks, a generator G and a discriminator D :

- **Generator**: produces synthetic data from a random $\mathbf{z} \sim p_Z$, where p_Z is a known distribution
- **Discriminator**: binary classifier, tries to distinguish real samples from fake ones

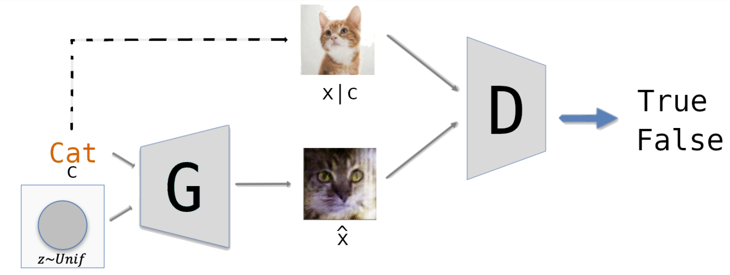
$$\min_G \max_D L(D, G) = \mathop{\mathbb{E}}_{\mathbf{x} \sim p_X} \left[\log(D(\mathbf{x})) \right] + \mathop{\mathbb{E}}_{\mathbf{z} \sim p_Z} \left[\log(1 - D(G(\mathbf{z}))) \right]$$



Conditional GAN

- Conditional variant of the GANs
- A constraint/label c is simply given as an input to both G and D
- Works well for generating image with a class constraints

$$\min_G \max_D L(D,G) = \mathop{\mathbb{E}}_{\substack{c\sim p_C \\ \mathbf{x}\sim p_{X|C}}} \Big[\log(D(\mathbf{x},c))\Big] + \mathop{\mathbb{E}}_{\substack{\mathbf{z}\sim p_Z \\ c'\sim p_C}} \Big[\log(1-D(G(\mathbf{z},c'),c'))\Big]$$



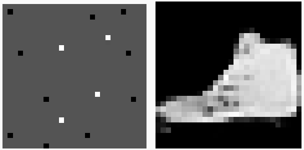
Question: Can GANs be conditioned in any way other than with labels ?

Question: Can GANs be conditioned in any way other than with labels ?

A solution: Auxiliary tasks

Contributions

- **Image reconstruction**
- **Polarimetric data generation**



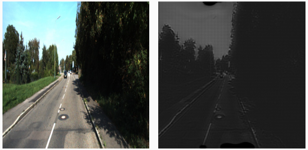


Image reconstruction

Context

Generative Adversarial Networks

Image reconstruction

Data augmentation of polarimetric datasets

Conclusion

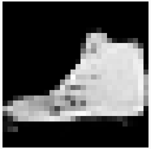
Problem

Objectives

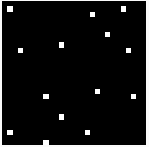
- Generation under pixel constraints
- Unstructured information

Differences with inpainting

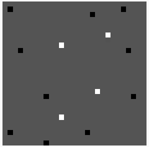
- Very few information ($\sim 0.5\%$)
- Full-size image generation



(a) Original Image



(b) Binary Mask



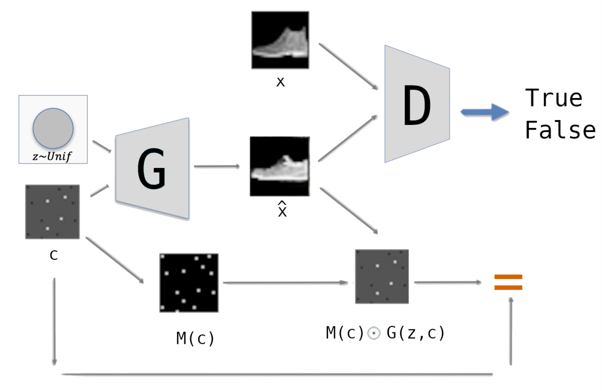
(c) Pixel Constraints

Constrained GAN

Theoretical objective

$$\min_G \max_D L(D,G) = \mathop{\mathbb{E}}_{\mathbf{x}\sim p_X}\Big[\log(D(\mathbf{x}))\Big] + \mathop{\mathbb{E}}_{\substack{\mathbf{z}\sim p_Z \\ c\sim p_C}}\Big[\log(1-D(G(\mathbf{z},c)))\Big] \quad \textbf{GAN task}$$

s.c. $c = M(c)\odot G(\mathbf{z},c)$, where $M(C)$ gives the binary mask of the constraints

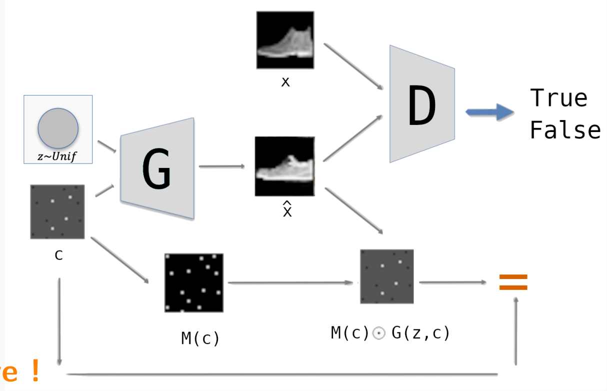


Constrained GAN

Theoretical objective

$$\min_{G}\max_{D} L(D,G) = \mathop{\mathbb{E}}_{\mathbf{x}\sim p_X}\left[\log(D(\mathbf{x}))\right] + \mathop{\mathbb{E}}_{\substack{\mathbf{z}\sim p_Z\\ c\sim p_C}}\left[\log(1-D(G(\mathbf{z},c)))\right] \quad \textbf{GAN task}$$

s.c. $c = M(c)\odot G(\mathbf{z},c)$, where $M(C)$ gives the binary mask of the constraints



Difficult to solve !

Relaxation of the constrained problem

Reconstruction:

$$c = M(c) \odot G(\mathbf{z}, c) + E$$

where $E \in \mathbb{R}^{n \times m}$ is a random noise (model error).

Maximum A Posteriori:

$$\begin{aligned} \mathbf{x}^* &= \arg\max_{\mathbf{x}} \log p_{X|C}(\mathbf{x}|c) \\ &= \arg\max_{\mathbf{x}} \log p_X(\mathbf{x}) + \log p_{C|X}(c|\mathbf{x})(+const) \end{aligned}$$

Parametrized MAP: $\mathbf{x} \to G(\mathbf{z}, c)$

$$G^* = \arg\max_{G} \mathop{\mathbb{E}}_{\substack{c \sim p_C \\ \mathbf{z} \sim p_Z}} \left[\log p_X(G(\mathbf{z}, c)) + \log p_{C|X}(C|G(\mathbf{z}, c)) \right]$$

Relaxation of the constrained CGAN

Recall $$G^* = \arg\max_G \mathop{\mathbb{E}}_{\substack{c\sim p_C \\ \mathbf{z}\sim p_Z}} \left[\log p_X(G(\mathbf{z},c)) + \log p_{C|X}(C|G(\mathbf{z},c))\right]$$

Assumption: $E \sim \mathcal{N}[0,\Sigma^2]$, for $c = M(c) \odot G(\mathbf{z},c) + E$. Then

$$\mathop{\mathbb{E}}_{\substack{c\sim p_C \\ \mathbf{z}\sim p_Z}} \log p_{C|X}(C|G(\mathbf{z},C))] = \mathop{\mathbb{E}}_{\substack{\mathbf{z}\sim p_Z \\ c\sim p_C}} \left[\|c - M(c)\odot G(\mathbf{z},C)\|_2^2\right]$$

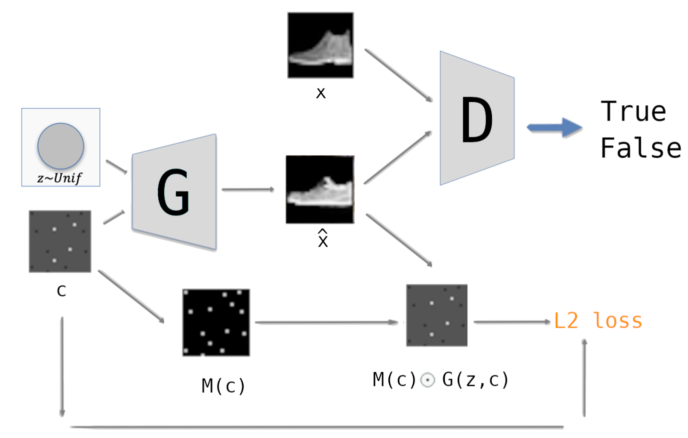
Final problem:

$$\min_G \max_D L_{reg}(D,G) = L(D,G) + \lambda\, \mathbb{E}_{\substack{\mathbf{z}\sim p_Z \\ c\sim p_C}} \left[\|c - M(c)\odot G(\mathbf{z},c)\|_2^2\right]$$

Relaxation of the constrained CGAN

Final problem:

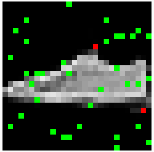
$$\min_G \max_D L_{reg}(D,G) = L(D,G) + \lambda \mathop{\mathbb{E}}_{\substack{\mathbf{z}\sim p_Z \\ c\sim p_C}} \left[\|c - M(c) \odot G(\mathbf{z}, c)\|_2^2 \right]$$



Evaluation

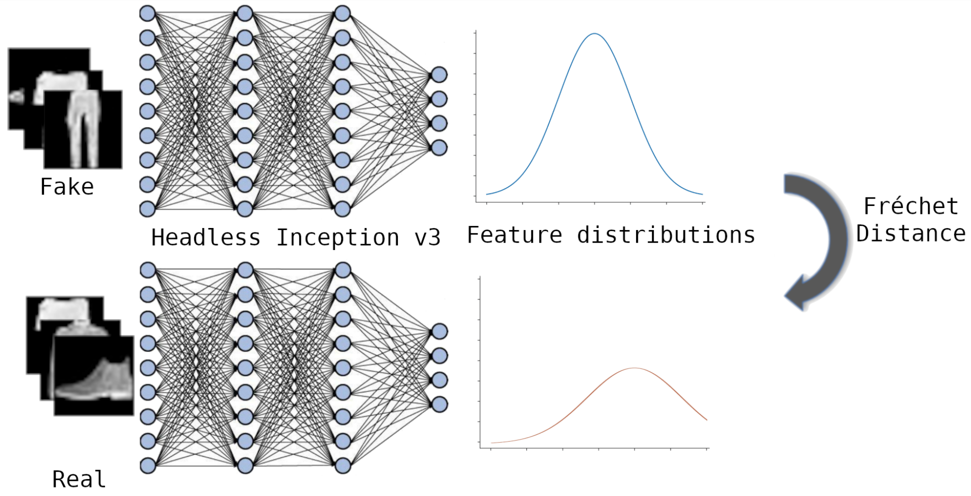
Metrics

- Respect of the constraints: Mean Square Error on constrained pixels
- Visual quality: No explicit way of measuring !
  Solution: Fréchet Inception Distance



Martin Heusel et al., "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", NeurIPS2017

Fréchet Inception Distance



Martin Heusel et al., "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", NeurIPS2017

Experimental setting

Task: Hyperparameter search on λ

- Objective: find evidence of a controllable trade-off between quality and respect of the constraints
- FashionMNIST dataset, 10% of the set used to sample constraints

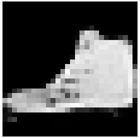
Networks architecture: DCGAN-like[1]

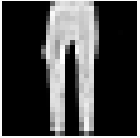
- Very small networks
- Generator: 2 deconvolutional layers + 1 dense
- Discriminator: 2 convolutional layers + 1 dense

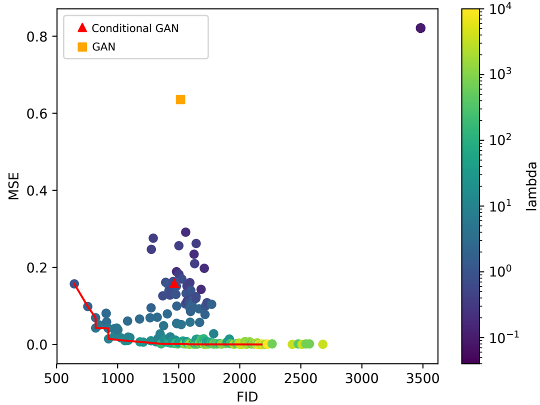
[1] Alec Radford et al., "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", ICLR2016

Results on FashionMNIST

- Trade-off clearly visible
- Adding constraints can enhance visual quality
- Reconstruction task enhances both quality and respect of constraints

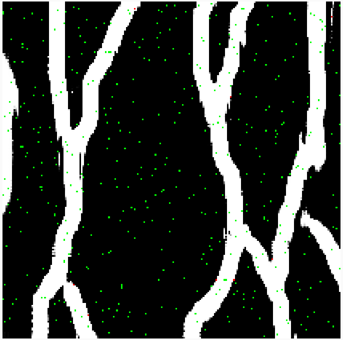






Practical application: underground terrain dataset

- Collaboration with SCK.CEN (Belgium)
- Hard patterns to learn, higher dimension (200x200)
- Can be learned with fully convolutional GANs $\rightarrow$ fewer parameters, shorter training time



Data augmentation of polarimetric datasets

Context

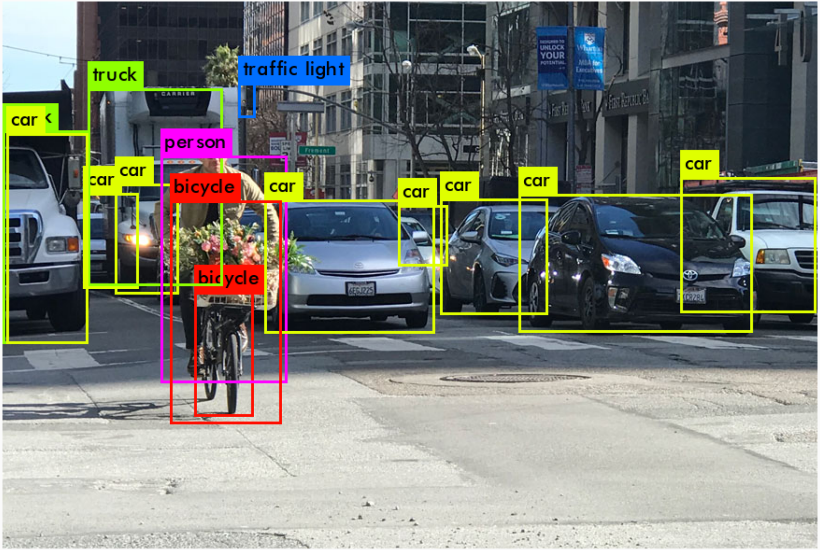
Generative Adversarial Networks

Image reconstruction

Data augmentation of polarimetric datasets

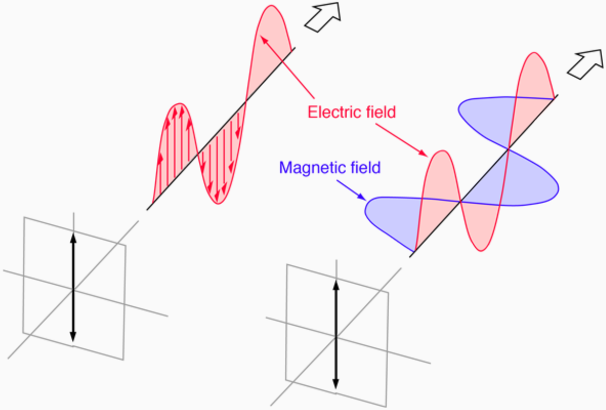
Conclusion

Object detection in road scenes

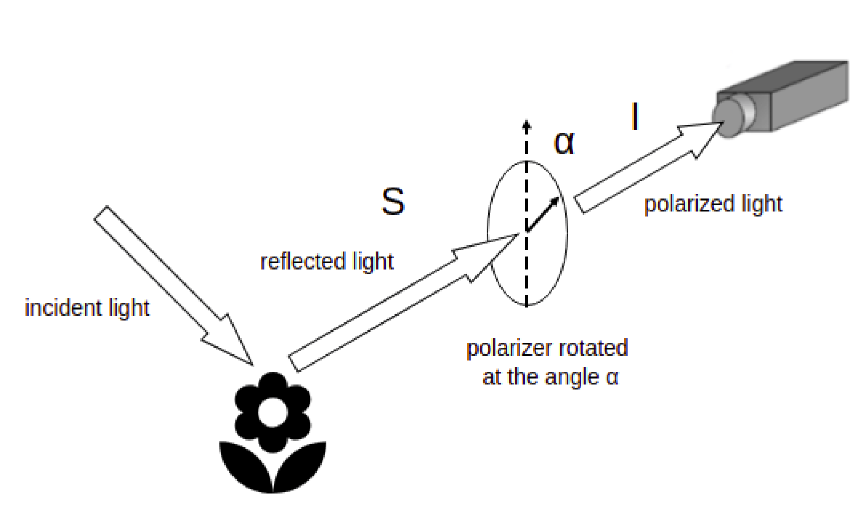


Bad weather conditions

Polarized light



Polarimetric camera



Polarimetric imaging

Polarimetric image:

I_0

I_{45}

I_{90}

I_{135}

Polarimetric imaging

Polarimetric image:

I_0

I_{45}

I_{90}

I_{135}

Problem: no large labeled datasets

PolarLITIS: 2,469 (paired) labeled images, BDD100K: 100,000 labeled RGB images

Polarimetric image generation

Image translation: RGB to polarimetric

Objective

Data augmentation in the polarimetric domain (very few existing large-size datasets)

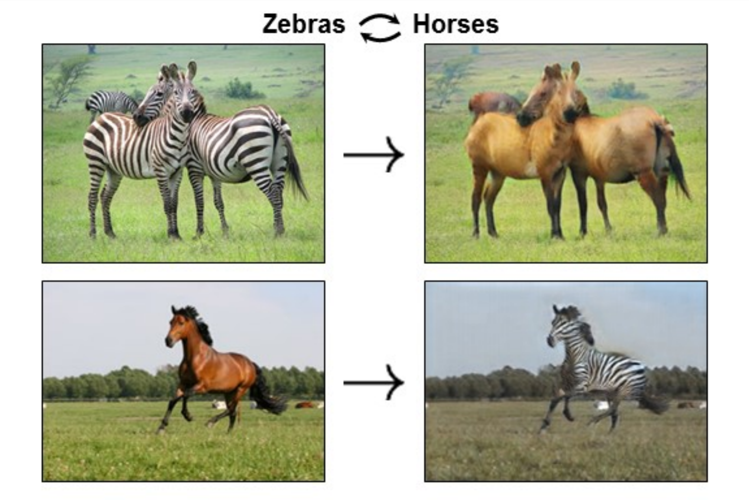
Polarimetric image generation

Image translation: RGB to polarimetric

Objective
Data augmentation in the polarimetric domain (very few existing large-size datasets)

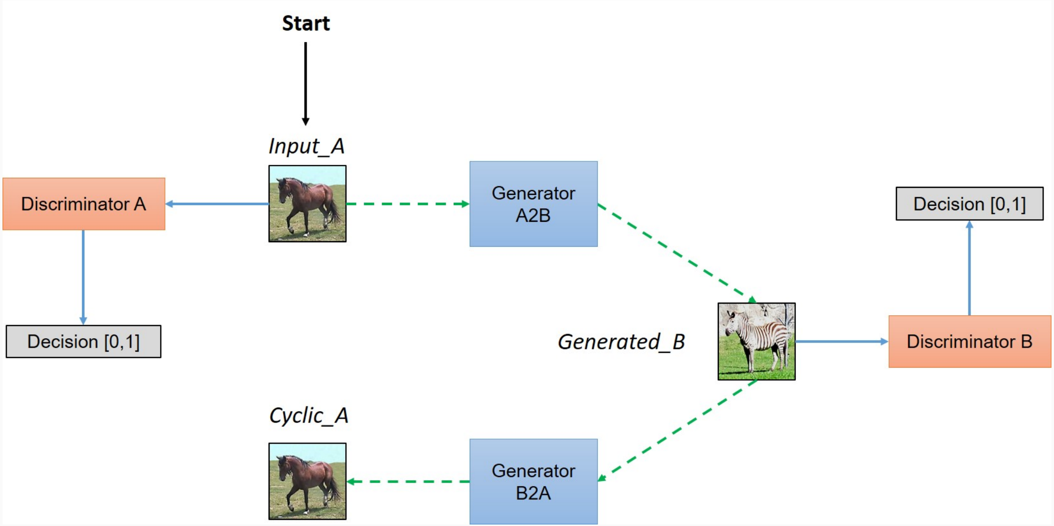
Ill-posed problem !

Domain transfer: a first case of auxiliary task



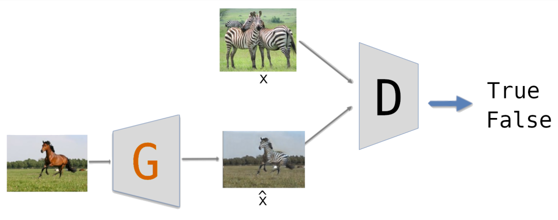
No paired data !

Domain transfer

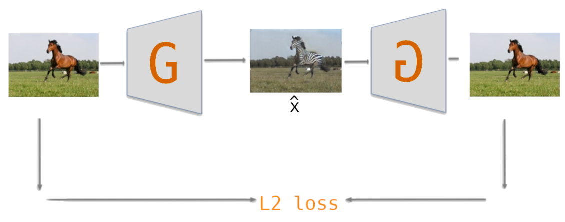


CycleGAN

Adversarial task



Cyclic consistency: reconstruction task



Properties of polarimetric images

Polarimetric image: 1 channel per polarizer angle

$$\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_{45}, \mathbf{y}_{90}, \mathbf{y}_{135}]^\top$$

Stokes parameters Vector that describe the polarization state in terms of intensity ($\mathbf{s}_0$) and degrees of polarization ($\mathbf{s}_1$, $\mathbf{s}_2$)

$$\mathbf{s} = [\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2]^\top$$

Computing the Stokes parameters

$$\mathbf{y}_{i,j} = \mathbf{A}\mathbf{s}_{i,j}, \forall i \leq n, j \leq p$$
$$\mathbf{s}_{i,j} = f(\mathbf{y}_{i,j}, \mathbf{A}), \forall i \leq n, j \leq p$$

where $\mathbf{A}$ is the *calibration matrix*, constant and unique to the camera.

Calibration constraint

Problem 1: A is not invertible for all cameras

In our works, we use a non square calibration matrix

$$\mathbf{A} = \frac{1}{2}\begin{bmatrix} 1 & \cos(2\alpha_1) & \sin(2\alpha_1) \\ 1 & \cos(2\alpha_2) & \sin(2\alpha_2) \\ 1 & \cos(2\alpha_3) & \sin(2\alpha_3) \\ 1 & \cos(2\alpha_4) & \sin(2\alpha_4) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

Problem 2: hard constraints on s

$$\mathbf{s}_0^2 \geq \mathbf{s}_1^2 + \mathbf{s}_2^2$$

Squared sum of polarized intensities cannot exceed squared total light intensity

Proposed solutions

Solution to problem 1: use $\mathbf{A}^\dagger$ is the pseudo-inverse of $\mathbf{A}$ as

$$\hat{\mathbf{s}}_{i,j} = \mathbf{A}^\dagger \mathbf{y}_{i,j} \forall i \leq n, j \leq p \ .$$

Property

$\mathbf{y}_{i,j} = \mathbf{A}\mathbf{A}^\dagger \mathbf{y}_{i,j}$ is satisfied iff $\mathbf{y}_{i,j} \in \ker(\mathbf{A}\mathbf{A}^\dagger - Id)$.
For our specific calibration matrix $\mathbf{A}$ the solution to this constraint is

$$\left\{ \mathbf{y} = \begin{bmatrix} \mathbf{y}_0 & \mathbf{y}_{45} & \mathbf{y}_{90} & \mathbf{y}_{135} \end{bmatrix}^\top \middle| \mathbf{y}_0 + \mathbf{y}_{90} = \mathbf{y}_{45} + \mathbf{y}_{135} \right\} \ .$$

Calibration auxiliary task: Relaxation of the constraint

$$L_{calib}(\mathbf{y}) = \mathop{\mathbb{E}}_{y \sim p_Y} ||\mathbf{y}_{i,j} - \mathbf{A}\mathbf{A}^\dagger \mathbf{y}_{i,j}||^2$$

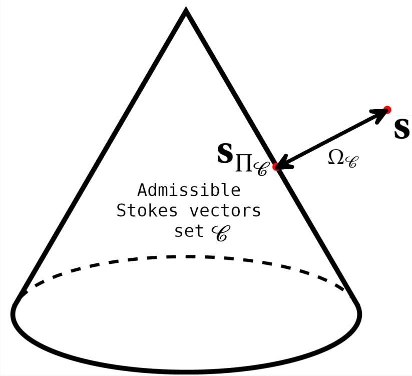
Proposed solutions

Solution for problem 2: Optical admissibility auxiliary task. Since

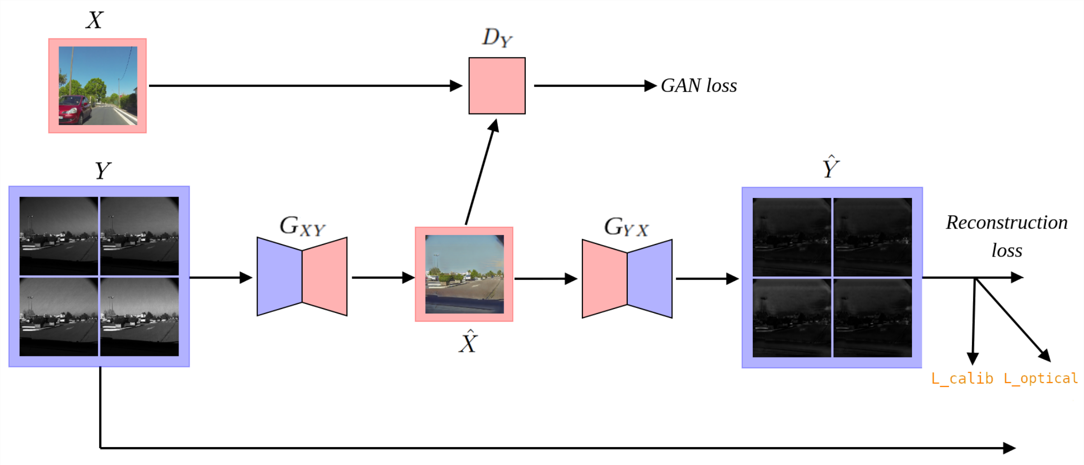
$$\mathbf{s}_0^2 \geq \mathbf{s}_1^2 + \mathbf{s}_2^2 \ ,$$

we can relax this constraint by minimizing

$$L_{optical}(\mathbf{s}) = \mathop{\mathbb{E}}_{y \sim p_Y} \max\left(\hat{\mathbf{s}_1}^2 + \hat{\mathbf{s}_2}^2 - \hat{\mathbf{s}_0}^2, 0\right)$$

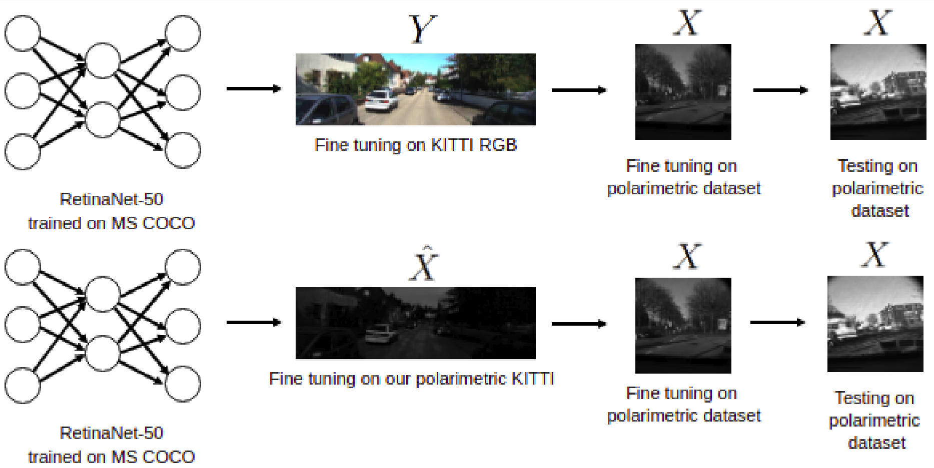


Our approach

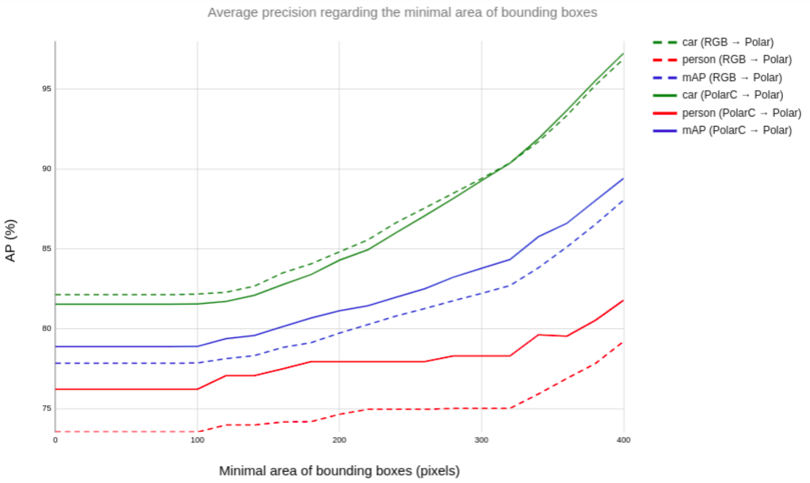


Visual results

Quantitative evaluation: object detection task



Experimental evaluation



Average precision up to 9% improvement in object detection on polarimetric images

Perspectives: projection operator

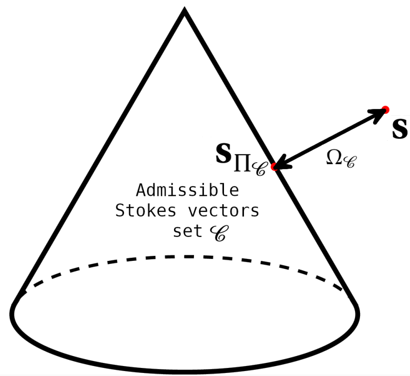
Optical constraint as a set of admissible solutions

Recall the optical constraint

$$\mathbf{s}_0^2 \geq \mathbf{s}_1^2 + \mathbf{s}_2^2$$

Can be formulated as a set of solutions

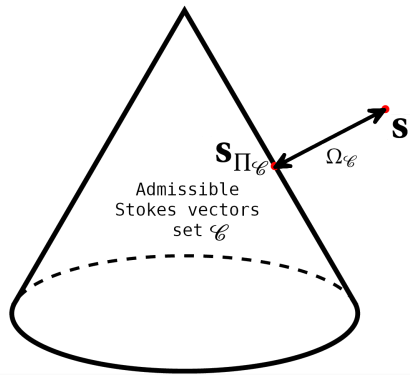
$$\mathcal{C} = \left\{ (\mathbf{s}_0, \mathbf{s}_{1,2}) \in \mathbb{S} \;\middle|\; \|\mathbf{s}_{1,2}\|_2 \leq \mathbf{s}_0,\ \mathbf{s}_{1,2} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} \right\} ,$$



Perspectives: projection operator

Closed-form projection operator

$$\Pi_{\mathcal{C}}(\mathbf{s}_0, \mathbf{s}_{1,2}) = \begin{cases} (\mathbf{s}_0, \mathbf{s}_{1,2}) & \text{if} \quad \|\mathbf{s}_{1,2}\|_2 \leq \mathbf{s}_0 \\ \frac{1+\mathbf{s}_0/\|\mathbf{s}_{1,2}\|_2}{2}(\|\mathbf{s}_{1,2}\|_2, \mathbf{s}_{1,2}) & \text{if} \quad \|\mathbf{s}_{1,2}\|_2 > \mathbf{s}_0 \end{cases}$$



Two different approaches

- Project the output of the GAN $(\hat{\mathbf{y}} = G(\mathbf{x}) \to \hat{\mathbf{y}} = \mathbf{A}\Pi_{\mathcal{C}}\mathbf{A}^{\dagger}G(\mathbf{x}))$
- Optimize a proximal distance

Perspectives: projection operator

Proximal distance

$$\Omega_{\mathcal{C}} = \|\mathbf{A}^{\dagger} G(\mathbf{y}) - \Pi_{\mathcal{C}}(\mathbf{A}^{\dagger} G(\mathbf{y}))\|^2$$

with closed-form gradient

$$\nabla_G \Omega_{\mathcal{C}}(\mathbf{s}) = (\mathbf{s} - \Pi_{\mathcal{C}}(\mathbf{s})) \times \begin{cases} 0 & \text{if} \quad \|\mathbf{s}_{1,2}\|_2 \leq \mathbf{s}_0 \\ \nabla_G \mathbf{s} - \nabla_G \frac{1}{2}\left[(1 + \frac{\mathbf{s}_0}{\|\mathbf{s}_{1,2}\|_2})(\|\mathbf{s}_{1,2}\|_2, \mathbf{s}_{1,2})\right] & \text{if} \quad \|\mathbf{s}_{1,2}\|_2 > \mathbf{s}_0 \end{cases}$$

Conclusion

Context

Generative Adversarial Networks

Image reconstruction

Data augmentation of polarimetric datasets

Conclusion

Conclusion

Auxiliary tasks

- Powerful approach for conditioning of GANs
- Leverage on domain-specific knowledge instead of labeled data
- Adapted to different kind of conditioning (equality, inequality, set membership, ...)

Main contributions

Contributions in image reconstruction

- Proposed a controllable approach for image reconstruction with very few pixels
- Highlighted a trade-off between quality and conditioning of the images
- Applied it to image reconstruction and geology

Main contributions

Contributions in polarimetric image generations

- Proposed a set of constraints for generating polarimetric images
- Proposed an approach for transferring color images to the polarimetric domain
- Showed that generated polarimetric images enhance performances of detection models
- Produced a polarimetric version of BDD100K and KITTI

Perspectives

Image reconstruction

- More work on the architectures
- Extending the approach to different prior distributions
- Extending the approach to different settings

Polarimetric image generation

- Enhance the visual quality of smaller objects with better architectures
- Experiment with the projection operator

- **Dilated Spatial Generative Adversarial Networks for Ergodic Image Generation**. **Cyprien Ruffino**, Romain Hérault, Eric Laloy, and Gilles Gasso, In: CAp 2017
- **Pixel-Wise Conditioning of Generative Adversarial Networks**. **Cyprien Ruffino**, Romain Hérault, Eric Laloy, and Gilles Gasso, In: ESANN 2019
- **Pixel-Wise Conditioned Generative Adversarial Networks for Image Synthesis and Completion**. **Cyprien Ruffino**, Romain Hérault, Eric Laloy, and Gilles Gasso, In: Neurocomputing
- **Generating Polarimetric-Encoded Images Using Constrained Cycle-Consistent Generative Adversarial Networks**. Rachel Blin*, **Cyprien Ruffino***, Samia Ainouz, Romain Hérault, Gilles Gasso, Fabrice Mériaudeau, and Stéphane Canu, In: Currently in Preparation
- **Gradient-Based Deterministic Inversion of Geophysical Data with Generative Adversarial Networks: Is It Feasible?**. Eric Laloy, Niklas Linde, **Cyprien Ruffino**, Romain Hérault, Gilles Gasso, and Diederik Jacques, In: Computers and Geosciences

Thank you for your attention ! :)

Appendix

Noises and distances

Reconstruction error Recall that we want to get $\max_{\mathbf{x}} \log p_{C|X}(c|x)$. We had

$$\hat{E} = c - M(c) \odot G(z, c) \quad \text{(We initially assumed } E \sim p_E \ (p_E = \mathcal{N}(0, \Sigma^2)))$$

As a distance: $L_{rec} = Div(p_{\hat{E}} || p_E)$

Exponential family of distributions: $p_\psi(\mathbf{x}, \theta) = h(\mathbf{x}) \exp(\mathbf{x}^\top \theta - \psi(\theta) - g_\psi(\mathbf{x}))$ [2]

Distribution	Distance measure
Gaussian	Squared Euclidian Distance
Multinomial	Kullback-Leibler divergence
Exponential	Itakura-Saito distance
Poisson	Relative entropy

[2] ψ, h and g_ψ convex Legendre function

Noises and distances

Generalized distance measures with Bregman divergence

$$\hat{E} = c - M(c) \odot G(z, c) \quad \text{(We initially assumed } E \sim p_E \; (p_E = \mathcal{N}(0, \Sigma^2)))$$

As a distance: $L_{rec} = Div(p_{\hat{E}} || p_E)$

We can instantiate this with a Bregman divergence

$$L_{rec} = D_\phi\Big(p_{\hat{E}} || p_E\Big) \quad ,$$

with $\phi(\theta)$ a convex function on θ the parameters of the distribution p_E.

Bregman divergences for exponential family distributions

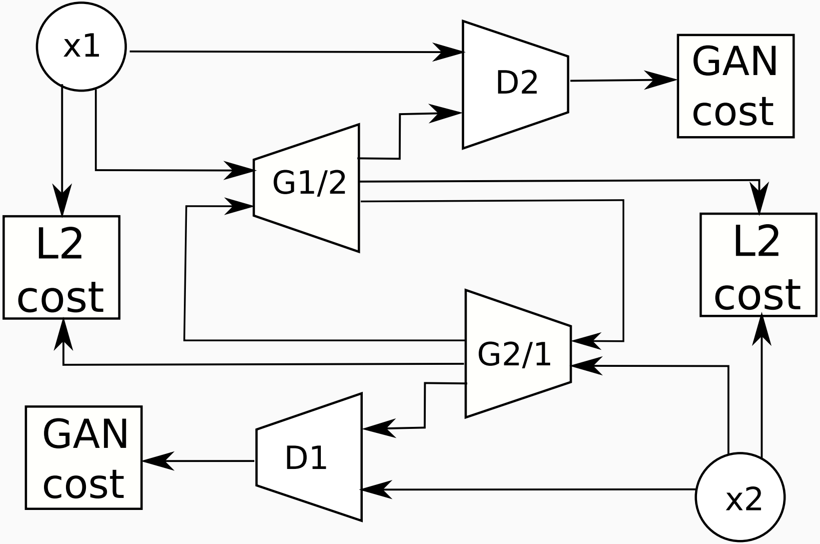
Distribution p_E	$\phi(\theta)$	$D_\phi(p_{\hat{E}} \| p_E)$
1D Gaussian	$\frac{1}{2\sigma^2}\mu^2$	$\mathbb{E}_{\epsilon \sim p_{\hat{E}}} \frac{1}{2\sigma^2}(\epsilon - \mu)^2$
1D Poisson	$\lambda \log \lambda - \lambda$	$\mathbb{E}_{\epsilon \sim p_{\hat{E}}} \epsilon \log(\frac{\epsilon}{\lambda}) - \epsilon + \lambda$
1D Bernoulli	$q \log q + (1-q)\log(1-q)$	$\mathbb{E}_{\epsilon \sim p_{\hat{E}}} \epsilon \log(\frac{\epsilon}{q}) + (1-\epsilon)\log(\frac{1-\epsilon}{1-q})$
1D Binomial	$Nq \log(\frac{Nq}{N}) + (N - Nq)\log(\frac{N-Nq}{N})$	$\mathbb{E}_{\epsilon \sim p_{\hat{E}}} \epsilon \log(\frac{\epsilon}{Nq}) + (N-\epsilon)\log(\frac{N-\epsilon}{N-Nq})$
1D Exponential	$-\ln(1/\lambda) - 1$	$\mathbb{E}_{\epsilon \sim p_{\hat{E}}} \frac{\epsilon}{1/\lambda} - \ln(\frac{\epsilon}{1/\lambda}) - 1$
dD Gaussian	$\frac{1}{(2\sigma^2)}\|\mu\|^2$	$\mathbb{E}_{E \sim p_{\hat{E}}} \frac{1}{2\sigma^2}\|E - \mu\|^2$
dD Multinomial	$\sum_{j=1}^{d} N_j q_j \log(\frac{N_j q_j}{N})$	$\mathbb{E}_{E \sim p_{\hat{E}}} \sum_{j=1}^{d} E_j \log(\frac{E_j}{N_j q_j})$

Arindam Banerjee et al., "Clustering with Bregman Divergences", JMLR 2005

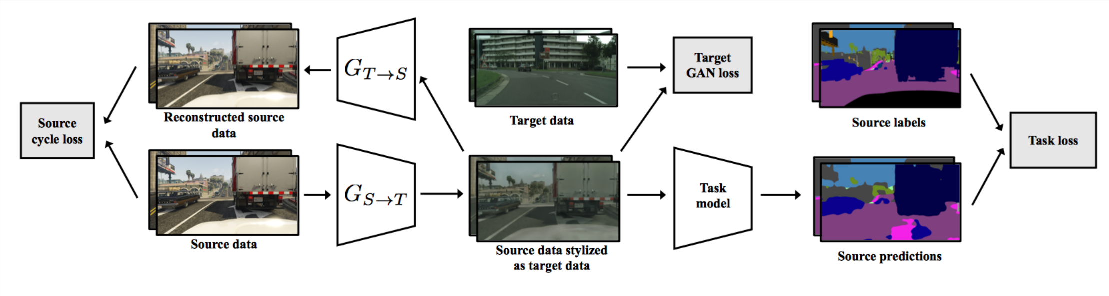
Evolution of the visual quality

2014 2015 2016 2017

Full CycleGAN



Generic auxiliary loss with task model



Judy Hoffman et al., "CyCADA: Cycle-consistent adversarial domain adaptation", ICML2018