Data Augmentation in High Dimensional Low Sample Size Setting with Geometry-Aware Variational Autoencoders

Clément Chadebec* & Elina Thibeau-Sutre†

*Université de Paris - INRIA (HeKA team) - INSERM
†Equipe Aramis - Institut du Cerveau

May 30, 2021

Overview

Main Challenges

Main challenges with medical data

- Small data sets:
  - potential poor subject variability
  - no statistically significant results
  - overfitting
- Large data (e.g. fMRI) $\Longrightarrow$ thousands of dimensions

Need for

- Data augmentation
- Dimensionality reduction

A solution ?

- Variational Autoecoders

Issue

- Unable to generate faithfully with small data sets

Classic Data Augmentation

- Adding some geometric transformations (shift, rotations ...)
- Adding noise, blur ...

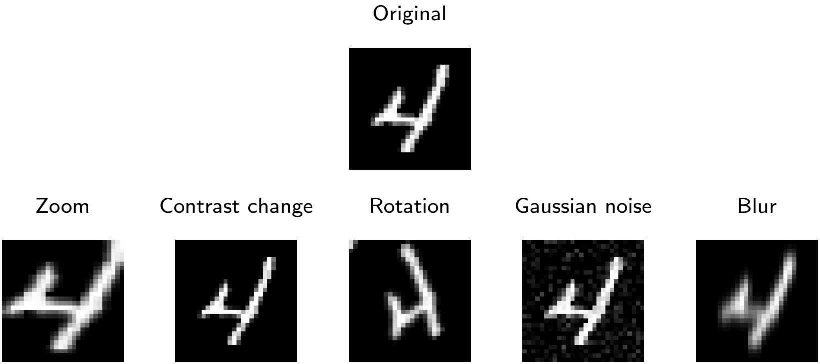


Figure: Examples of transformations

Classic Data Augmentation - Shortcomings

Classic DA

- Is data set dependent
- May require the intervention of an expert “knowledge”

Figure: Nine figure rotated.

An attractive solution ?

- Generative models (Generative Adversarial Networks, Variational Auto-Encoders ...)

Use of Generative Models for DA

GANs have already seen a wide use in many fields of application including medicine [YWB19]:

- Magnetic Resonance Images (MRI) [STR$^+$18, CMST17]
- Computed Tomography (CT) [FADK$^+$18, SYPS19]
- X-ray [MMKSM18, SVD$^+$18, WGG$^+$20],
- Positron Emission Tomography (PET) [BKK$^+$17],
- Mass spectroscopy data [LZL$^+$19],
- Dermoscopy [BAN18]
- Mammography [KRO$^+$18, WWCL18]

$\Longrightarrow$ Most of these studies involved either a quite large training set (above 1000 training samples) or quite small dimensional data.

$\Longrightarrow$ As of today, the HDLSS setting remains poorly explored.

$\Longrightarrow$ Use VAEs!

VAE - The Idea

- An auto-encoder based model...

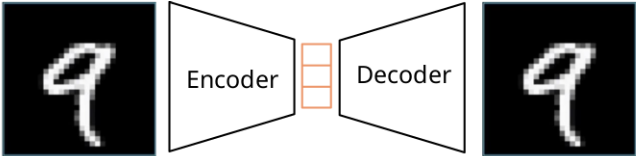


Figure: Simple Auto-Encoder

- ... but where an input data point is encoded as a **distribution** defined over the latent space [KW14, RMW14]

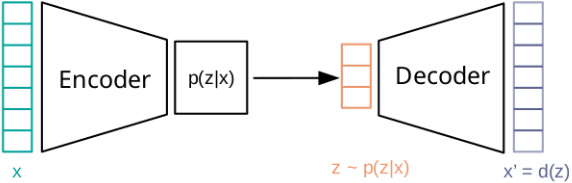


Figure: VAE framework

VAE - Mathematical Considerations

- Let $x \in \mathcal{X}$ be a set of data and $\{P_\theta, \theta \in \Theta\}$ a parametric model
- We assume there exists latent variables $z \in \mathcal{Z}$ living in a smaller space such that the marginal likelihood writes

$$p_\theta(x) = \int p_\theta(x|z) q_{\mathrm{prior}}(z) dz \,,$$

where q_{prior} is a prior distribution over the latent variables and $p_\theta(x|z)$ is referred to as the decoder

$$q_{\mathrm{prior}} = \mathcal{N}(0, I), \qquad p_\theta(x|z) = \prod_{i=1}^{D} \mathcal{B}(\pi_{\theta_i(z)})$$

Objective:

- Maximizing the likelihood of the model

Problem:

- The integral is often intractable making $p_\theta(z|x) = \frac{p_\theta(x|z) q_{\mathrm{prior}}(z)}{p_\theta(x)}$ intractable $\Longrightarrow$ Bayesian Inference is unusable

The ELBO

- We have to use Variational Inference

$$q_\phi(z|x) \simeq p_\theta(z|x)\,,$$

where $q_\phi(z|x) = \mathcal{N}(\mu_\phi(x), \Sigma_\phi(x))$

- This leads to an unbiased estimate of the log-likelihood

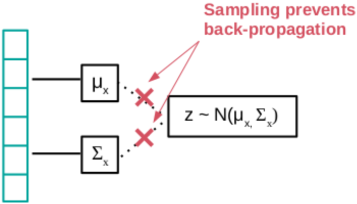
$$\widehat{p}_\theta(x) = \frac{p_\theta(x,z)}{q_\phi(z|x)}, \qquad \mathbb{E}_{z\sim q_\phi(z|x)}[\widehat{p}_\theta(x)] = p_\theta(x)\,,$$

- Taking the logarithm of the expectation we have

$$\begin{aligned}
\log p_\theta(x) &= \log \mathbb{E}_{z\sim q_\phi(z|x)}[\widehat{p}_\theta(x)] \\
&\geq \mathbb{E}_{z\sim q_\phi(z|x)}[\log(\widehat{p}_\theta(x))] \\
&\geq \mathbb{E}_{z\sim q_\phi(z|x)}[\log(p_\theta(x,z)) - \log(q_\phi(z|x))] \\
&\geq ELBO
\end{aligned}$$

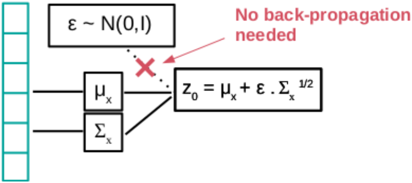
The Reparametrization Trick

- Since $z \sim \mathcal{N}(\mu_\phi(x), \Sigma_\phi(x))$, the model is not amenable to gradient descent



(a) Back-propagation impossible

$\Longrightarrow$



(b) Back-propagation possible

$\Longrightarrow$ Optimization with respect to encoder and decoder parameters made possible !

Generating new samples

- We only need to sample $z \sim \mathcal{N}(0, I)$ and feed it to the decoder.

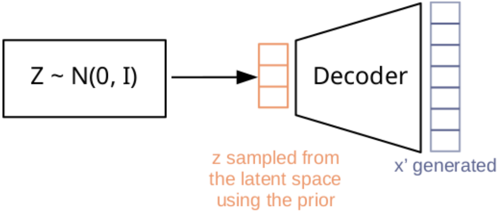


Figure: Generation procedure

Pros:

- Very simple to use in practice

Cons:

- The prior and posterior are not expressive enough to capture complex distributions
- Poor latent space prospecting

Defining a New Framework

Assumptions:

- As of now the latent space structure was supposed to be Euclidean (i.e. $\mathcal{Z} = \mathbb{R}^d$)
- Let us now relax this hypothesis and assume that $\mathcal{Z}$ is a Riemannian manifold endowed with a metric $\mathbf{G}$.
- It was shown that exploiting the geometrical aspect of probability distributions can lead to far more efficient sampling [GCC09, GC11]

Our ideas:

1. Exploit the manifold structure of the latent space to improve the posterior sampling [CMA20]
2. Learn the metric defined in the latent space [CMA20]
3. Use the learned geometry to generate instead of the prior [CTSBA21]

1) Improve Posterior Sampling - Riemannian HMC

- The idea relies on the Riemannian Hamiltonian Monte Carlo Sampler [GC11]
- Simulates the evolution $(z(t), v(t))$ of a particle whose motion is governed by Hamiltonian dynamics and having a potential $U(z)$ and kinetic energy $K(v, z)$

$$U(z) = -\log p_{\mathrm{target}}(z), \qquad K(v, z) = \frac{1}{2} v^\top \mathbf{G}^{-1}(z) v \,.$$

- Use of the "Generalized" Leapfrog integrator to sample from p_{target}
- The target density p_{target} is proportional to the true posterior:

$$p_\theta(z|x) = \frac{p_\theta(x, z)}{p_\theta(x)} \propto p_\theta(x, z) = p(x|z)p(z) = p_{\mathrm{target}}(z)\,.$$

Pros:

- Posterior sampling is guided by the gradient of the true posterior
- Use the underlying geometry of the data to improve sampling

Cons:

- The metric is unknown

2) Learn the Metric - The Choice of the Metric

- We propose to parametrize the metric as follows [Lou19]:

$$\mathbf{G}^{-1}(z) = \sum_{i=1}^{N} L_{\psi_i} L_{\psi_i}^{\top} \exp\Big(- \frac{\|z - c_i\|_2^2}{T^2} \Big) + \lambda I_d \,,$$

- L_{ψ_i} are lower triangular matrices parametrized using neural networks
- T is a temperature to smooth the metric
- c_i are the centroids
- λ is a regularization factor

Pros:

- Closed-form expression of the inverse metric $\Longrightarrow$ useful for geodesic computation
- Metric volume element $\sqrt{\det \mathbf{G}(z)}$ easily scalable through $\lambda \Longrightarrow$ geodesics travel through most populated areas

The Model - Riemannian Hamiltonian VAE

- The graphical scheme

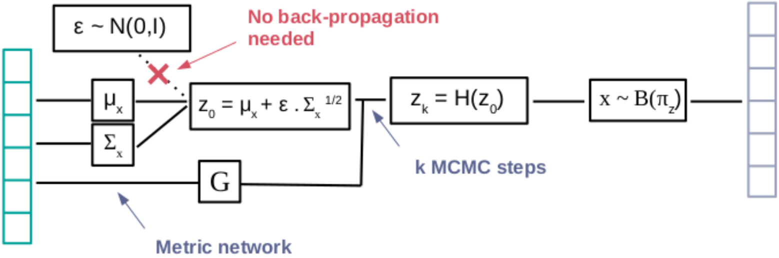
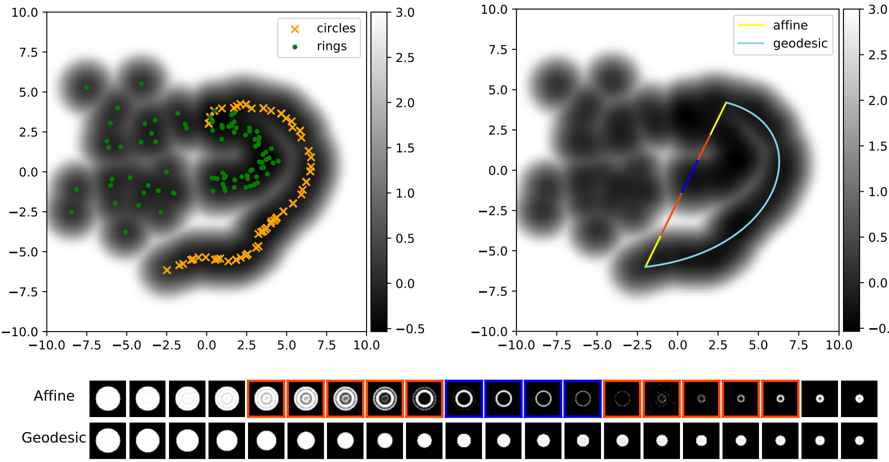
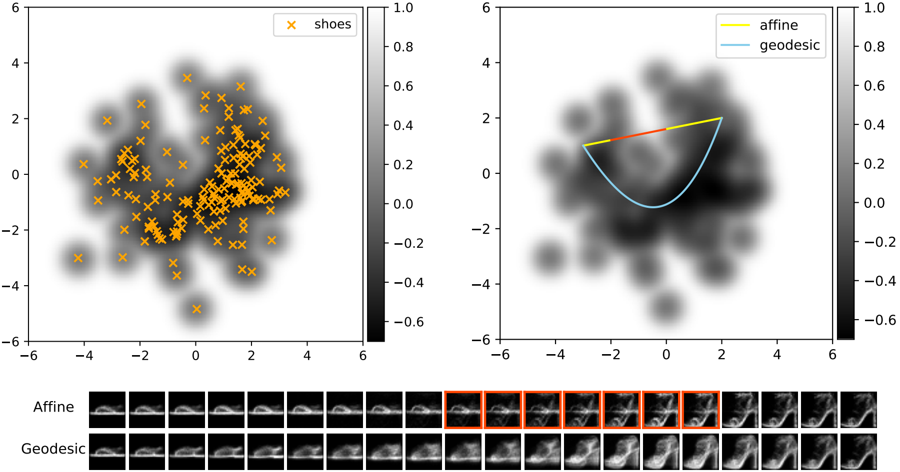


Figure: Riemannian Hamiltonian VAE.

The Learned Latent Space



The Learned Latent Space



3) Improve Data Generation - Sample With the Metric

Idea:

- Our idea is to use a geometry-based sampling procedure

$$p(z) = \frac{\rho_S(z)\sqrt{\det \mathbf{G}^{-1}(z)}}{\int\limits_{\mathbb{R}^d} \rho_S(z)\sqrt{\det \mathbf{G}^{-1}(z)dz}},$$

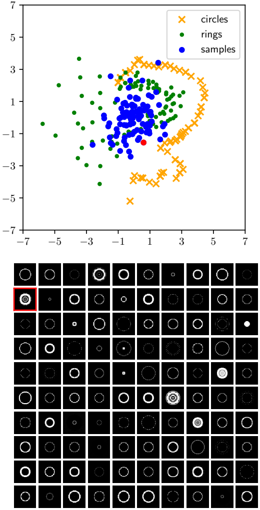
  where S is a compact set and $\rho_S(z) = 1$ if $z \in S$, 0 otherwise.
- Use of classic MCMC sampler (e.g. Hamiltonian Monte Carlo)

Pros:

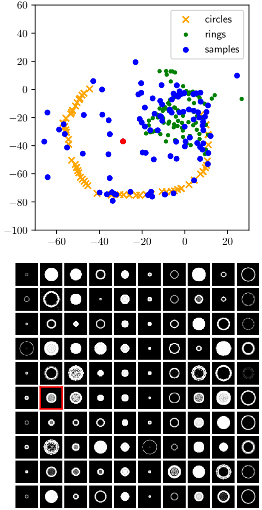
- $\mathbf{G}^{-1}$ easily computable
- Samples "close" to the data

Sampling Comparison

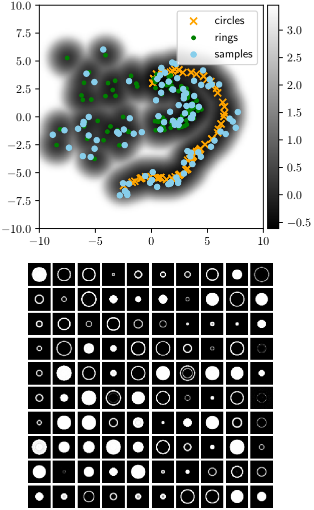
(a) VAE - $\mathcal{N}(0, I)$



(b) VAE - VAMP (multimodal)

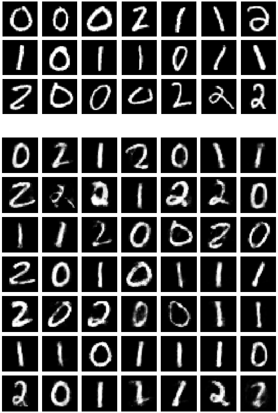


(c) Ours

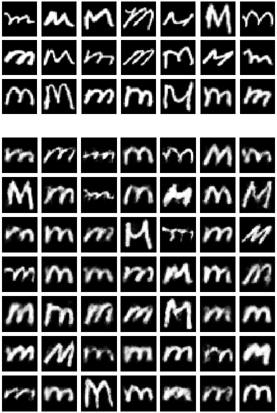


Sampling Comparison - Higher Dimension

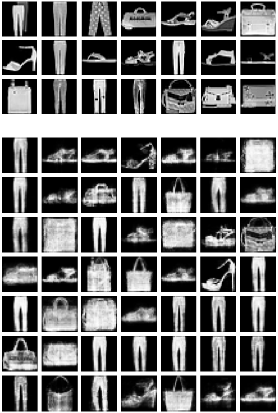
(a) *reduced* MNIST (120)



(b) *reduced* EMNIST (120)



(c) *reduced* Fashion (120)



Data Augmentation

Data Augmentation - Framework

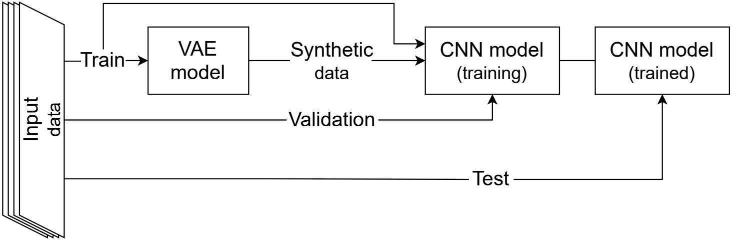


Figure: Data Augmentation framework

Data Augmentation

1. Toy Data
2. Medical Imaging

Robustness Across Data Sets

Table: Classification results on *reduced* data sets ($\sim$ 50 samples per class)

	MNIST	MNIST (unbal.)	EMNIST (unbal.)	FASHION
Baseline	89.9 ± 0.6	81.5 ± 0.7	82.6 ± 1.4	76.0 ± 1.5
Baseline + Synthetic				
Basic Augmentation (X5)	92.8 ± 0.4	86.5 ± 0.9	85.6 ± 1.3	77.5 ± 2.0
Basic Augmentation (X10)	88.2 ± 2.2	82.0 ± 2.4	85.7 ± 0.3	79.2 ± 0.6
Basic Augmentation (X15)	92.8 ± 0.7	85.8 ± 3.4	86.6 ± 0.8	80.0 ± 0.5
VAE - 200*	88.5 ± 0.9	84.0 ± 2.0	81.7 ± 3.0	78.6 ± 0.4
VAE - 2k*	92.2 ± 1.6	88.0 ± 2.2	86.0 ± 0.2	79.3 ± 1.1
Ours-200	91.0 ± 1.0	84.1 ± 2.0	85.1 ± 1.1	77.0 ± 0.8
Ours-500	92.3 ± 1.1	87.7 ± 0.9	85.1 ± 1.1	78.5 ± 0.9
Ours-1k	93.2 ± 0.8	$\mathbf{89.7 \pm 0.8}$	87.0 ± 1.0	$\mathbf{80.2 \pm 0.8}$
Ours-2k	$\mathbf{94.3 \pm 0.8}$	89.1 ± 1.9	$\mathbf{87.6 \pm 0.8}$	78.1 ± 1.8

* Using a standard normal prior to generate

- Classic DA is data set dependent
- Vanilla VAE performs as well as classic DA

Robustness Across Data Sets

Table: Classification results on *reduced* data sets ($\sim$ 50 samples per class) on synthetic samples only

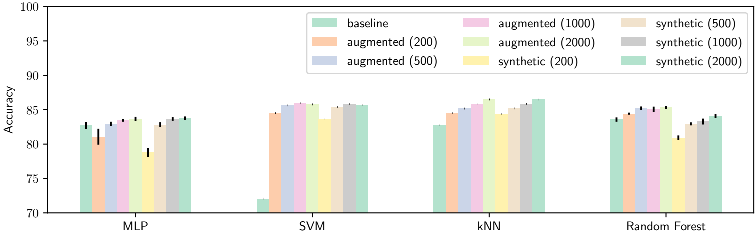
	MNIST	MNIST (unbal.)	EMNIST (unbal.)	FASHION
Baseline	89.9 ± 0.6	81.5 ± 0.7	82.6 ± 1.4	76.0 ± 1.5
Synthetic Only				
VAE - 200*	69.9 ± 1.5	64.6 ± 1.8	65.7 ± 2.6	73.9 ± 3.0
VAE - 2k*	86.5 ± 2.2	79.6 ± 3.8	78.8 ± 3.0	76.7 ± 1.6
Ours-200	87.2 ± 1.1	79.5 ± 1.6	77.0 ± 1.6	77.0 ± 0.8
Ours-500	89.1 ± 1.3	80.4 ± 2.1	80.2 ± 2.0	78.5 ± 0.8
Ours-1k	90.1 ± 1.4	86.2 ± 1.8	82.6 ± 1.3	79.3 ± 0.6
Ours-2k	92.6 ± 1.1	87.5 ± 1.3	86.0 ± 1.0	78.3 ± 0.9

* Using a standard normal prior to generate

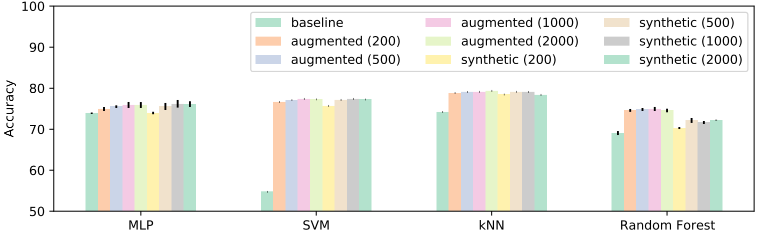
- The proposed model seems to create diverse samples relevant to the classifier

Robustness Across Classifiers

(a) *reduced* MNIST balanced



(b) *reduced* MNIST unbalanced



A Note on the Method Scalability

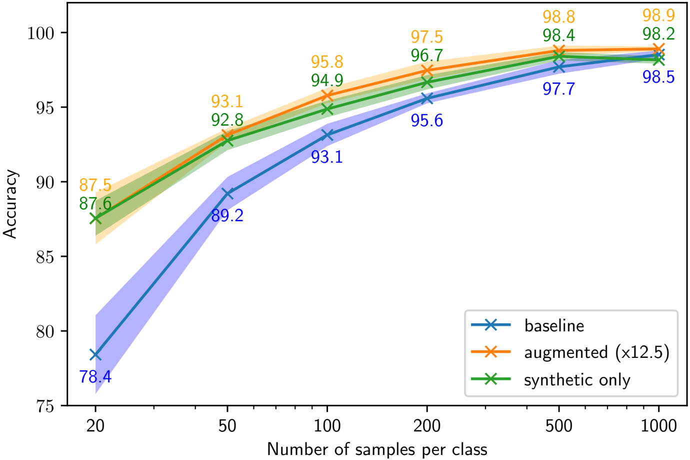


Figure: Benchmark classifier accuracy according to the number of samples in the training set on MNIST.

1. Toy Data
2. Medical Imaging

Datasets

Classification task: Alzheimer's disease patients (**AD**) vs Cognitively Normal participants (**CN**) using T1-weighted MR images.

Table: Summary of participant demographics, mini-mental state examination (MMSE) and global clinical dementia rating (CDR) scores at baseline.

Data set	Label	Obs.	Age	Sex M/F	MMSE	CDR
ADNI	CN	403	73.3 ± 6.0	185/218	29.1 ± 1.1	0: 403
	AD	362	74.9 ± 7.9	202/160	23.1 ± 2.1	0.5: 169, 1: 192, 2: 1
AIBL	CN	429	73.0 ± 6.2	183/246	28.8 ± 1.2	0: 406, 0.5: 22, 1: 1
	AD	76	74.4 ± 8.0	33/43	20.6 ± 5.5	0.5: 31, 1: 36, 2: 7, 3: 2

MRI preprocessing

Bias field correction (N4ITK) + linear registration (ANTS) + cropping

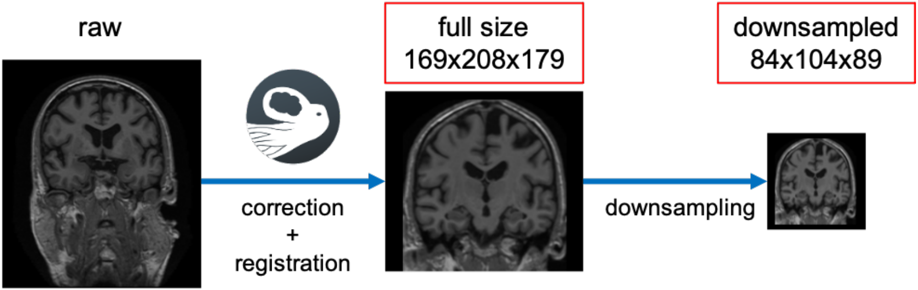
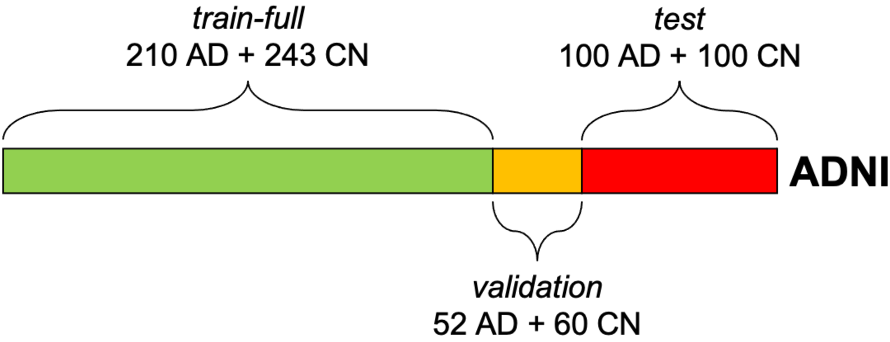


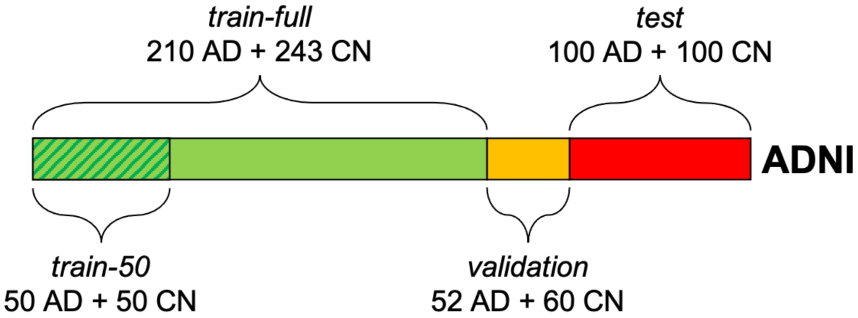
Figure: Preprocessed MRI used in the study

Find wonderful data at:
`/network/lustre/dtlake01/aramis/datasets/adni/caps/caps_v2021`

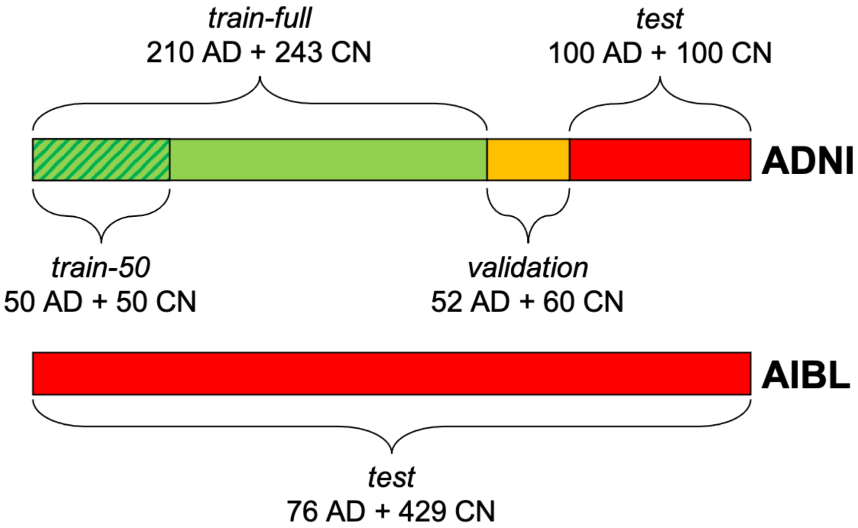
Evaluation procedure



Evaluation procedure

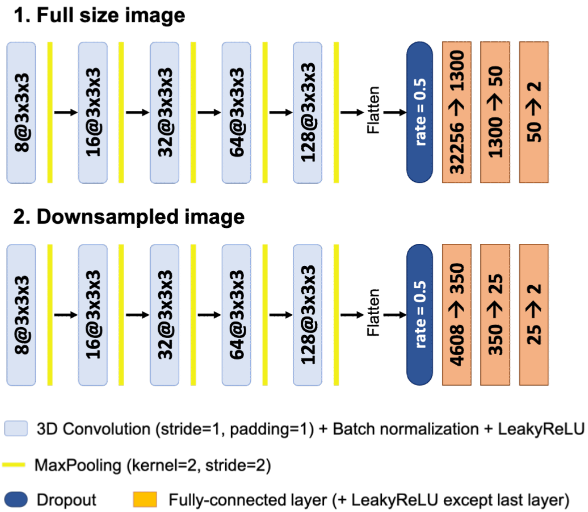


Evaluation procedure



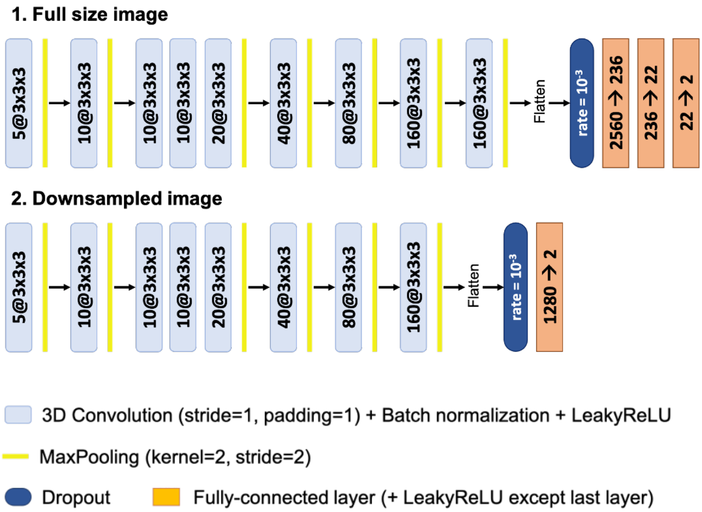
CNN architectures

Baseline architectures provided by a previous study [WTSDM+20]



CNN architectures

Optimized architectures found with random search procedure (ClinicaDL)



Experiments

Four series of experiments:

- **baseline** architecture on *train-50*
- **baseline** architecture on *train-full*
- **optimized** architecture on *train-50*
- **optimized** architecture on *train-full*

For each experiment 20 CNNs are run and the performance is the mean value of the 20 performance values.

Synthesized images

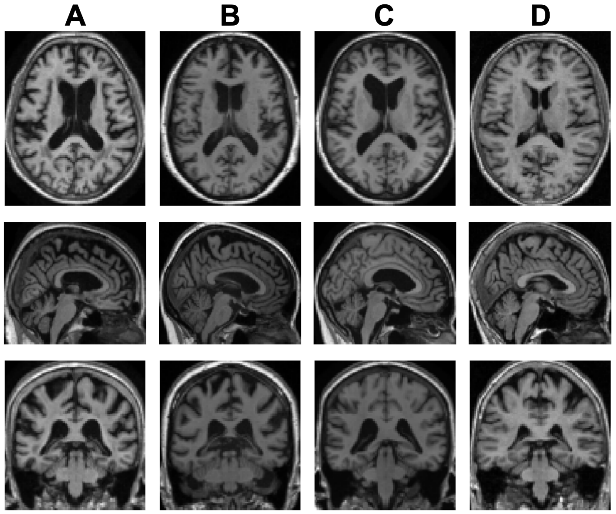


Figure: Example of two *true* patients compared to two generated by our method. Can you find the intruders ?

Synthesized images

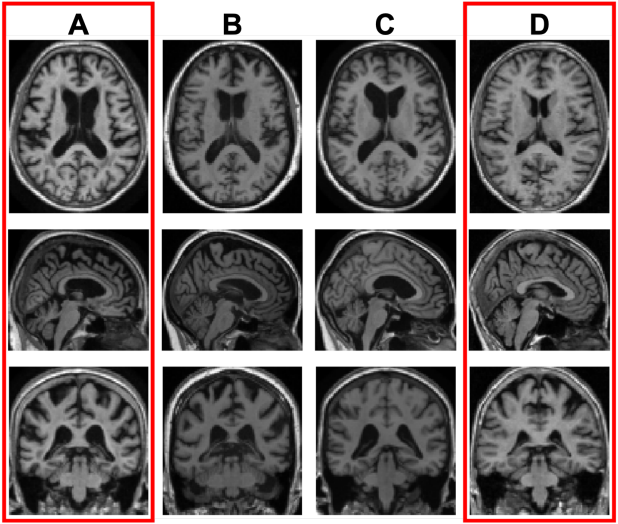


Figure: Example of two *true* patients compared to two generated by our method. Can you find the intruders ?

Results on train-50 with baseline CNN

Table: Mean test performance of each series of 20 runs trained with the **baseline** hyperparameters on *train-50* set.

data set	ADNI balanced accuracy	AIBL balanced accuracy
real	66.3 ± 2.4	67.2 ± 4.1
real (high-resolution)	67.9 ± 2.3	66.5 ± 3.0
500 synthetic + real	69.4 ± 1.6	68.5 ± 2.5
1000 synthetic + real	70.5 ± 2.1	70.6 ± 3.1
2000 synthetic + real	71.2 ± 1.6	72.8 ± 2.2
3000 synthetic + real	72.6 ± 1.6	73.6 ± 3.0
5000 synthetic + real	$\mathbf{74.1 \pm 2.2}$	$\mathbf{76.1 \pm 3.6}$
10000 synthetic + real	74.0 ± 2.7	74.9 ± 3.2

Increase of balanced accuracy of 6.2 points on ADNI and 8.9 points on AIBL

Results on train-full with baseline CNN

Table: Mean test performance of each series of 20 runs trained with the **baseline** hyperparameters on *train-full* set.

data set	ADNI balanced accuracy	AIBL balanced accuracy
real	77.7 ± 2.5	78.4 ± 2.4
real (high-resolution)	80.6 ± 1.1	80.4 ± 2.6
500 synthetic + real	82.2 ± 2.4	82.9 ± 2.5
1000 synthetic + real	84.4 ± 1.8	83.7 ± 2.3
2000 synthetic + real	85.9 ± 1.6	83.8 ± 2.2
3000 synthetic + real	85.8 ± 1.7	84.4 ± 1.8
5000 synthetic + real	85.7 ± 2.1	84.2 ± 2.2
10000 synthetic + real	$\mathbf{86.3 \pm 1.8}$	$\mathbf{85.1 \pm 1.9}$

Increase of balanced accuracy of 5.7 points on ADNI and 4.7 on AIBL

Results on train-50 with optimized CNN

Table: Mean test performance of each series of 20 runs trained with the **optimized** hyperparameters on *train-50* set.

data set	ADNI balanced accuracy	AIBL balanced accuracy
real	75.5 ± 2.7	75.6 ± 4.1
real (high-resolution)	72.1 ± 3.1	71.2 ± 5.1
500 synthetic + real	75.6 ± 2.5	76.0 ± 4.2
1000 synthetic + real	77.8 ± 2.3	80.9 ± 3.2
2000 synthetic + real	76.9 ± 2.4	80.0 ± 3.6
3000 synthetic + real	77.8 ± 1.9	81.2 ± 3.7
5000 synthetic + real	76.9 ± 2.5	80.9 ± 2.7
10000 synthetic + real	**78.0±2.1**	**81.9±2.2**

Increase of balanced accuracy of 2.5 points on ADNI and 6.3 points on AIBL

Results on train-full with optimized CNN

Table: Mean test performance of each series of 20 runs trained with the **optimized** hyperparameters on *train-full* set.

data set	ADNI balanced accuracy	AIBL balanced accuracy
real	85.5 ± 2.4	81.9 ± 3.2
real (high-resolution)	85.7 ± 2.5	84.4 ± 1.7
500 synthetic + real	86.0 ± 1.8	83.2 ± 2.4
1000 synthetic + real	86.5 ± 1.9	83.7 ± 2.0
2000 synthetic + real	$\mathbf{87.2 \pm 1.7}$	84.0 ± 2.0
3000 synthetic + real	85.8 ± 2.6	83.6 ± 3.2
5000 synthetic + real	86.4 ± 1.3	83.5 ± 2.2
10000 synthetic + real	86.7 ± 1.8	$\mathbf{84.3 \pm 1.8}$

Increase of balanced accuracy of 1.5 point on ADNI and -0.1 point on AIBL

Conclusion

Validation of a new VAE-based data augmentation framework on classification tasks on *toy* and *real-life* data sets.

Strengths:

- **Data set generalization** from 2D images (MNIST, EMNIST, FASHION) to 3D medical images (ADNI and AIBL),
- **Classifier independence** MLP, random forest, k-NN and SVM (on toy data sets) ; baseline and optimized parameters (on medical images).
- **Synthetic data relevance** classifiers achieved a similar or better classification performance when trained only on synthetic data than on the *real* train set.
- **Low sample size data sets usability** adding synthetic data improves classification performance even with a small training set (*train-50*)

Conclusion

Validation of a new VAE-based data augmentation framework on classification tasks on *toy* and *real-life* data sets.

Limitations - what could be improved:

- no extensive search on VAE hyperparameters.
- can it be easily coupled with other techniques to limit overfitting?
- would it benefit from the use of longitudinal data?
- *train-50* is still large compared to some medical data sets...

The End

Thank you !

Appendices

Clustering

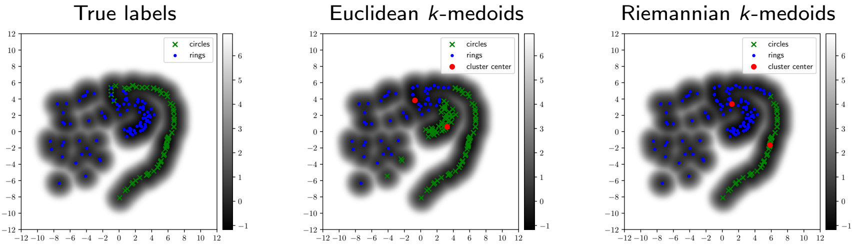


Figure: Euclidean and Riemannian k-medoids custering.

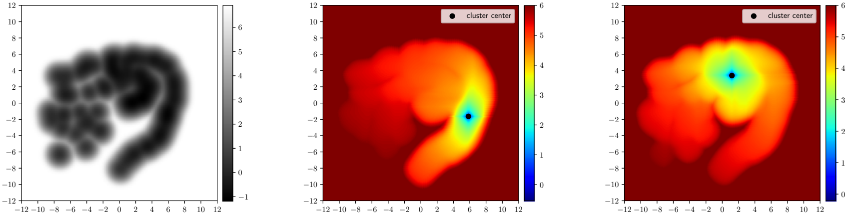


Figure: Distance maps.

Results - Clustering

Data set	Model	Subset 1	Subset 2	Subset 3	Mean
Synthetic data	linear	53.88	62.52	71.63	62.68
	geodesic	**71.41**	**81.39**	**79.49**	**77.43**
MNIST 1	linear	89.73	93.11	91.80	91.55
	geodesic	**91.68**	**94.51**	**95.63**	**93.94**
MNIST 2	linear	68.24	69.22	79.05	71.17
	geodesic	**70.35**	**71.34**	**79.64**	**73.78**
MNIST 3	linear	75.55	75.76	81.70	77.67
	geodesic	**76.08**	**77.94**	**81.96**	**78.66**
FashionMNIST 1	linear	90.47	91.63	86.78	89.63
	geodesic	**91.44**	**92.55**	**87.46**	**90.48**
FashionMNIST 2	linear	92.20	91.26	93.30	92.25
	geodesic	**93.56**	**91.80**	**94.12**	**93.16**
FashionMNIST 3	linear	72.46	79.58	83.16	78.40
	geodesic	**74.89**	**81.88**	**84.83**	**80.53**

Table: F1-Scores.

Tweaking the Approximate Posterior

- The ELBO can written as

$$ELBO = \log p_\theta(x) - \underbrace{\mathrm{KL}(q_\phi(z|x)||p_\theta(z|x))}_{\approx 0 \text{ if } q_\phi(z|x) \approx p_\theta(z|x)} .$$

- Since the Kullback-Leiber divergence is always non-negative, the objective is to try to make it vanish by tweaking the approximate posterior $q_\phi(z|x)$
- The idea is to add some Markov Chain Monte Carlo steps targeting the true posterior $p_\theta(z|x)$[SKW15]
- How to ensure that the model would still be amenable to the back-propagation ?

Normalizing Flows

- The idea is to use smooth invertible parametrized mappings f_ψ to "sample" z [RM15]
- K transformations are then applied to a latent variable z_0 drawn from an initial distribution q (here $q = q_\phi$) leading to a final random variable $z_K = f_x^K \circ \cdots \circ f_x^1(z_0)$ whose density writes

$$q_\phi(z_K|x) = q_\phi(z_0|x) \prod_{k=1}^{K} |\det \mathrm{J}_{f_x^k}|^{-1}, \tag{1}$$

Riemannian Hamiltonian VAE

- The idea relies on the Riemannian Hamiltonian Monte Carlo Sampler [GC11]
- We define a target density π:

$$p_\theta(x|z) = \frac{p_\theta(x,z)}{p_\theta(x)} \propto p_\theta(x,z) = \pi_x(z)\,.$$

- An auxiliary position-specific random variable $\rho \sim \mathcal{N}(0, \mathbf{G}(z))$ is introduced, the "momentum"
- The Hamiltonian writes

$$H_x^{Riem}(z,\rho) = U_x(z) + \frac{1}{2}\log((2\pi)^D \det \mathbf{G}(z)) + \frac{1}{2}\rho^\top \mathbf{G}(z)^{-1}\rho\,.$$

$\Longrightarrow$ Make use of the "Generalized" Leapfrog integrator

Pros:

- The sampling is guided by the gradient of the true posterior

References I

Christoph Baur, Shadi Albarqouni, and Nassir Navab, *Generating highly realistic images of skin lesions with GANs*, OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Springer, 2018, pp. 260–267.

Lei Bi, Jinman Kim, Ashnil Kumar, Dagan Feng, and Michael Fulham, *Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs)*, Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment, LNCS, Springer, 2017, pp. 43–51.

Clément Chadebec, Clément Mantoux, and Stéphanie Allassonnière, *Geometry-aware hamiltonian variational auto-encoder*, arXiv:2010.11518 [cs, math, stat] (2020).

Francesco Calimeri, Aldo Marzullo, Claudio Stamile, and Giorgio Terracina, *Biomedical data augmentation using generative adversarial neural networks*, International conference on artificial neural networks, Springer, 2017, pp. 626–634.

References II

Clément Chadebec, Elina Thibeau-Sutre, Ninon Burgos, and Stéphanie Allassonnière, *Data augmentation in high dimensional low sample size setting using a geometry-based variational autoencoder*, arXiv preprint arXiv:2105.00026 (2021).

Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan, *GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification*, Neurocomputing **321** (2018), 321–331.

Mark Girolami and Ben Calderhead, *Riemann manifold langevin and hamiltonian monte carlo methods*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73** (2011), no. 2, 123–214.

Mark Girolami, Ben Calderhead, and Siu A Chin, *Riemannian manifold hamiltonian monte carlo*, arXiv preprint arXiv:0907.1100 (2009).

References III

Dimitrios Korkinof, Tobias Rijken, Michael O'Neill, Joseph Yearsley, Hugh Harvey, and Ben Glocker, *High-resolution mammogram synthesis using progressive generative adversarial networks*, arXiv preprint arXiv:1807.03401 (2018).

Diederik P. Kingma and Max Welling, *Auto-encoding variational bayes*, arXiv:1312.6114 [cs, stat] (2014).

Maxime Louis, *Computational and statistical methods for trajectory analysis in a Riemannian geometry setting*, PhD Thesis, Sorbonnes universités, 2019.

Yufei Liu, Yuan Zhou, Xin Liu, Fang Dong, Chang Wang, and Zihong Wang, *Wasserstein gan-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology*, Engineering **5** (2019), no. 1, 156–163.

References IV

Ali Madani, Mehdi Moradi, Alexandros Karargyris, and Tanveer Syeda-Mahmood, *Chest x-ray generation and data augmentation for cardiovascular abnormality classification*, Medical Imaging 2018: Image Processing, vol. 10574, International Society for Optics and Photonics, 2018, p. 105741M.

Danilo Rezende and Shakir Mohamed, *Variational inference with normalizing flows*, International Conference on Machine Learning, PMLR, 2015, pp. 1530–1538.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra, *Stochastic backpropagation and approximate inference in deep generative models*, International conference on machine learning, PMLR, 2014, pp. 1278–1286.

Tim Salimans, Diederik Kingma, and Max Welling, *Markov chain monte carlo and variational inference: Bridging the gap*, International Conference on Machine Learning, 2015, pp. 1218–1226.

References V

Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine P Andriole, and Mark Michalski, *Medical image synthesis for data augmentation and anonymization using generative adversarial networks*, International Workshop on Simulation and Synthesis in Medical Imaging, LNCS, Springer, 2018, pp. 1–11.

Hojjat Salehinejad, Shahrokh Valaee, Tim Dowdell, Errol Colak, and Joseph Barfett, *Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks*, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 990–994.

Veit Sandfort, Ke Yan, Perry J. Pickhardt, and Ronald M. Summers, *Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks*, Scientific reports **9** (2019), no. 1, 16884.

- Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman, and Plácido Rogerio Pinheiro, *Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection*, Ieee Access **8** (2020), 91916–91923.
- Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, and Olivier Colliot, *Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation*, Medical Image Analysis **63** (2020), 101694.
- Eric Wu, Kevin Wu, David Cox, and William Lotter, *Conditional infilling gans for data augmentation in mammogram classification*, Image analysis for moving organ, breast, and thoracic images, Springer, 2018, pp. 98–106.
- Xin Yi, Ekta Walia, and Paul Babyn, *Generative adversarial network in medical imaging: A review*, Medical image analysis **58** (2019), 101552.