Continuous Methods : Hamiltonian Domain Translation

Emmanuel Menier, Michele Alessandro Bucci, Mouadh Yagoubi, Lionel Mathelin, Marc Schoenauer

Continuous Methods : Hamiltonian Domain Translation

Emmanuel Menier [1,2] **Michele Alessandro Bucci** [1] **Mouadh Yagoubi** [2] **Lionel Mathelin** [3] **Marc Schoenauer** [1]

Abstract

This paper proposes a novel approach to domain translation. Leveraging established parallels between generative models and dynamical systems, we propose a reformulation of the Cycle-GAN architecture. By embedding our model with a Hamiltonian structure, we obtain a continuous, expressive and most importantly invertible generative model for domain translation.

1. Introduction

Domain translation is the process of transforming elements from one domain to another. One can think of applications such as neural style transfer (Gatys et al., 2016) which is for example used to apply a certain painter's style to photo-realistic images. A common problem encountered in domain translation applications is that, in many cases, paired data is not available during training, which means that the problem has to be formulated in an unsupervised setting. Unsupervised learning is very common in the field of generative modeling, and several architectures have been proposed to deal with the problem of Unsupervised Domain Translation. In this work, we focus on the Cycle-GAN (Zhu et al., 2017) architecture, which has proved successful in various applications of Unsupervised Domain Translation [*].

Despite its success, the formulation of the Cycle-GAN method has been questioned and shown to be ill-posed. Using results from Chen & Gopinath (2000), it can be shown that when considering two distinct domains, there exist an infinity of pairings between the two domains which satisfy the Cycle-GAN objective. This is an issue as the model could get stuck trying to learn wildly inefficient mappings, leading to unsatisfactory optima. This conditioning problem has been explored in depth by de Bézenac et al. (2021), as they proposed to use a regularized residual network to learn the mapping between two given domains. Borrowing ideas from optimal transport and dynamical systems, they showed that pushing the training towards simple, low-energy, transformations in latent space leads to learning a sensible and trivially invertible mapping between the two domains of interest.

The study of the links between dynamical systems theory and deep learning is still to this day a major topic of interest. One can for example cite the identification of residual networks as first order approximations of a time-continuous process which has led to the development of ground-breaking approaches such as neural ordinary differential equations (Neural ODE (Chen et al., 2018)) or invertible neural networks (Behrmann et al., 2019).

Building on this existing connection, as well as the work of de Bezenac *et al.*, we propose a formulation of unsupervised domain translation as a continuous time process with conservation guarantees which ensure invertibility by construction. The proposed architecture learns the dynamics of the transformation as a Hamiltonian dynamical system. Hamiltonian systems are typically used in General Mechanics to describe the evolution of conservative systems. They preserve a quantity, called the Hamiltonian, along their trajectory. Using neural networks to learn Hamiltonian dynamics is an earlier idea that was proposed in Greydanus et al. (2019). However this work proposes to use them to ensure invertibility of the generative process which is a desirable property to ensure the domain translation problem is well-posed. Learning conservative transformations is in fact critical to other generative modeling approaches, such as normalizing flows (Rezende & Mohamed, 2015).

---

[1]TAU, Inria / LISN, Université Paris-Saclay / CNRS, Orsay, France [2]IRT - SystemX, Palaiseau, France [3]LISN, Université Paris-Saclay & CNRS, Orsay, France. Correspondence to: Emmanuel Menier <emmanuel.menier@inria.fr>.

[*]Cycle GAN project page, https://junyanz.github.io/CycleGAN/

2. Method

2.1. Invertibility and CycleGAN

Formally, we can look at the two domains as two separate sets $\mathcal{A}, \mathcal{B} \subset \mathbb{R}^d$, where d is the dimension of the space, *i.e.* the pixel space for images, or any latent representation space. The goal of unsupervised domain translation is to learn the forward mapping $F : \mathcal{A} \to \mathcal{B}$ as well as the reverse map $R : \mathcal{B} \to \mathcal{A}$ so that the pair (F, R) generates semantically meaningful samples of each domain. That is to say, the generated samples should be indistinguishable from samples in the target domain, while remaining coherent with their corresponding sample in the original domain.

CycleGan proposes to enforce these constraints by using a combined loss: $\mathcal{L} = \mathcal{L}_{adv} + \mathcal{L}_{cyc}$. The first term $\mathcal{L}_{adv} = \mathcal{D}(F(\mathcal{A}), \mathcal{B})$ corresponds to an adversarial loss which measures the distance between the generated samples and the target domain. This term ensures that generated samples are indistinguishable from the target domain. In CycleGAN, $\mathcal{D}$ is implemented using Generative Adversarial Networks (Goodfellow et al., 2014).

The second term in the loss is called the cyclic loss, $\mathcal{L}_{cyc} = \|F \circ R(x_\mathcal{A}) - x_\mathcal{A}\| + \|R \circ F(x_\mathcal{B}) - x_\mathcal{B}\|$. This term promotes transformations F that are invertible and such that $R = F^{-1}$. Intuitively, this pushes the CycleGAN architecture towards learning minimal transformations of the samples, so as to retain a maximum of information from the initial sample and simplify the reconstruction $R \circ F$. This second term is used to ensure coherence between the translated and initial samples. In addition, learning an invertible (thus bijective) map between the two domains is critical at the conceptual level. Indeed, one sample from a given domain should not map to multiple samples in the target domain as only one sample in the target domain should optimally satisfy the trade-off between coherence with the original sample and similarity with the target domain.

2.2. Continuous models and Hamiltonian Neural Networks

The previous paragraph outlined the importance of ensuring the translation map is invertible to relax the learning problem. In fact, this is not specific to the domain translation problem, as invertibility of learned maps has been linked to classical deep learning problems such as vanishing/exploding gradients in recurrent neural networks (Pascanu et al., 2012), or the training of other generative models like normalizing flows (Rezende & Mohamed, 2015). Several approaches have been proposed to push learned models towards invertibility (Miyato et al., 2018; Rosenblatt, 1952), however, they often impose significant constraints on the structure and expressivity of the models, leading to important training costs.

In this work, we propose to use a natural formulation for invertible transformations. Exploiting the parallel between the residual networks used in numerous image processing approaches, and ordinary differential equations, we propose to define domain translation as a continuous system. Starting at $t = 0$ with samples from one domain $x_{t=0} \in \mathcal{A}$, we learn a transport flow f_θ so that, at $t = T$, $x_{t=T} \in \mathcal{B}$:

$$\frac{dx}{dt} = f_\theta(x), \quad \text{s.t.} \quad x_0 \in \mathcal{A}, \ x_T \in \mathcal{B} \tag{1}$$

This formulation is not enough to ensure invertibility of the transformation, as the flow f_θ could be dissipative, or even unstable. To enforce invertibility, we express the flow f_θ as a conservative operator using Hamiltonian neural networks inspired from Greydanus et al. (2019). To do so, the samples are divided into two vectors of equal length $x = [p, q]$, (we assume d to be even as a modeling choice). In general mechanics, p and q would respectively describe the position and momentum of the studied entities. In our setting, their significance is more abstract and is defined by another function, called the Hamiltonian $\mathcal{H}_\theta(p, q) : \mathbb{R}^{d/2} \times \mathbb{R}^{d/2} \to \mathbb{R}$, which we parameterize using a neural network, hence:

$$f_\theta(x) = \begin{pmatrix} \frac{dp}{dt} & = & -\frac{\partial \mathcal{H}_\theta}{\partial q} \\ \frac{dq}{dt} & = & \frac{\partial \mathcal{H}_\theta}{\partial p} \end{pmatrix} \tag{2}$$

Using Neural ODEs and automatic differentiation, the function $\mathcal{H}_\theta$ can be trained to satisfy the transport objective, *i.e.* $x_T \in \mathcal{B}$ given $x_0 \in \mathcal{A}$. Moreover, this formulation is invertible by design as it preserves the quantity $\mathcal{H}_\theta(x)$ along its trajectory. We show below that learning the transformation f_θ with this formulation allows for the generation of semantically correct samples, without using the cyclic loss required in CycleGAN. Thanks to the conservation properties of the flow f_θ,

the inverse map is trivially obtained by integrating the flow backward in time:

$$x_{\mathcal{B}} = x_{\mathcal{A}} + \int_0^T f_\theta(x)dt \quad \Longleftrightarrow \quad x_{\mathcal{A}} = x_{\mathcal{B}} + \int_T^0 f_\theta(x)dt \tag{3}$$

3. Results

3.1. Generative results

We apply our Hamiltonian domain translation approach to image generation tasks. As proposed in de Bézenac et al. (2021), we train an encoder E and a decoder G to map images from both domains to a latent space of size $d = 128$. This is a common approach in many image processing approaches, as the intrinsic dimension of a given image processing problem is often much lower than the dimension of the pixel space. Thus, encoding images to a low-dimensional latent space reduces the domain translation problem complexity, as well as training costs.

Once the pair (E, G) is trained, it can be used to generate low-dimensional encoded vectors of images of the dataset at hand. We then use our approach to learn the transport flow f_θ. The Hamiltonian $\mathcal{H}$ and discriminator $\mathcal{D}$ are implemented as multi layer perceptron with 3 hidden layers. The continuous flow is learned using the *optimise-then-discretise* version of NeuralODEs. We apply the architecture to the task of translating male samples of the celebA (Liu et al., 2015) dataset to females. Figure 1 presents samples generated with this approach.
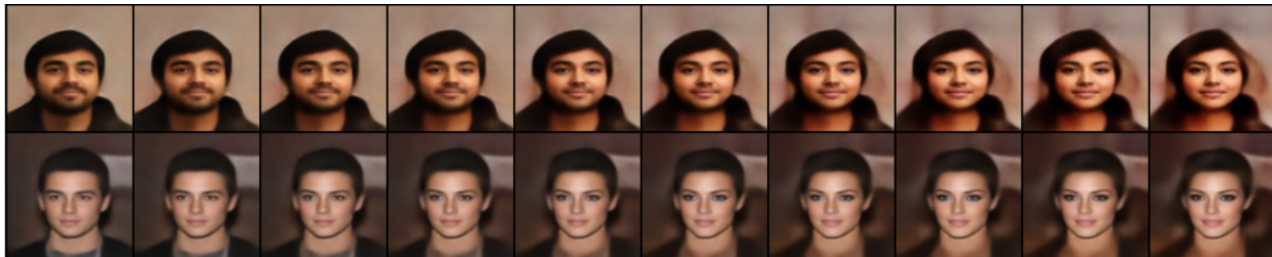

Figure 1. Selected samples of the male to female transport process using the proposed continuous domain translation approach. The transported encodings are decoded at regular time intervals, to illustrate the transformation applied by the model.

Figure 2. Selected samples of the reverse male to female generation process. NB : These samples are generated using equation (3) as transport flow was solely trained to map male to females.

As shown on Figure 1, decoding the transported samples along the transformation trajectory shows that the flow f_θ progressively transforms the male samples to females. As expected, the conservative nature of the model promotes transformations that retain non gender-specific features, as we observe that attributes such as pose, skin tone, face shape and background are preserved during the transformation. Figure 2 demonstrates an additional interest of the Hamiltonian architecture as we are able to generate males from female samples by simply integrating the flow backward (see Eq. 3). One should note that these results were obtained without ever training the model to map females to males as we do not compute the cyclic loss used in CycleGAN. These results are similar to the results of de Bézenac et al. (2021) while no penalization of the magnitude of the flow applied by the model is used but invertibility is enforced instead.

| T = 0.0 t.u. | T = 0.25 t.u. | T = 0.5 t.u. | T = 0.75 t.u. | T = 1.0 t.u. | T = 1.25 t.u. | T = 1.5 t.u. | T = 1.75 t.u. |

Figure 3. Results of excessive integration of the transport flow. A model trained to map males to females in 1 $t.u.$ is integrated backward for more than twice the map horizon. We observe that the generated samples retain semantic sense for up to about 1.5 times the training horizon.

3.2. Excessive Integration

An interesting feature of using a continuous flow to carry out domain translation is that one can gain some insight in the way the model transforms samples. If a flow f_θ has been trained to map two domains in one time unit (*t.u.*), $T = 1$, it can be integrated for a longer period, pushing the transformation further. This is one of the major differences between learning continuous transformations and discrete residual blocks. While residual blocks approximate the flow in specific regions of the latent space, the continuous flow is defined over the whole space. Any trained model starts losing performance once it drifts too far from its training conditions but we observed interesting results when integrating our model for several *t.u.*. Figure 3 shows that transported samples retain semantic meaning for up to about one and a half *t.u.*, as the model progressively adds more and more gender-related features such as beards, wider jaws, shorter hair, etc. This generalisation performance can be linked to the conservative architecture of the model which prevents it from diverging to unknown conditions. It also adds to the interest of the approach as it supports the idea that the model is consistent with the structure of the latent space.

3.3. Training

It should be noted that, once the autoencoder is trained, learning the flow f_θ is very inexpensive. The model starts generating semantically coherent samples after a single epoch, and does not require fine tuning between the training of the discriminator $\mathcal{D}$ and the flow f_θ. More formal benchmarking against comparable domain translation methods are planned for the future.

4. Conclusion

This work proposes a novel and improved formulation for domain translation. By using a time-continuous approach, we are able to leverage results from general mechanics to obtain a model that is invertible by construction. We show that this model can quickly learn to map two domains of interest, even in a latent space learned prior to training the domain translation architecture.

With the recent success of diffusion models (Nichol et al., 2021; Rombach et al., 2021), which are based on successive transformations in a pre-defined space, the analogy between generative models and dynamical systems becomes more and more relevant. In this context, the comparison of our proposed method with existing approaches such as normalizing flows and latent diffusion models constitutes the next step to this work, for instance exploring the potential extension of our continuous generative approach to stochastic differential equations (Li et al., 2020).

References

Behrmann, J., Grathwohl, W., Chen, R. T. Q., Duvenaud, D., and Jacobsen, J.-H. Invertible residual networks. In *ICML*, pp. 573–582, 2019. URL http://proceedings.mlr.press/v97/behrmann19a.html.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Chen, S. and Gopinath, R. Gaussianization. In Leen, T., Dietterich, T., and Tresp, V. (eds.), *Advances in Neural Information*

Processing Systems, volume 13. MIT Press, 2000. URL https://proceedings.neurips.cc/paper/2000/file/3c947bc2f7ff007b86a9428b74654de5-Paper.pdf.

de Bézenac, E., Ayed, I., and Gallinari, P. Cyclegan through the lens of (dynamical) optimal transport. In Oliver, N., Pérez-Cruz, F., Kramer, S., Read, J., and Lozano, J. A. (eds.), *Machine Learning and Knowledge Discovery in Databases. Research Track*, pp. 132–147, Cham, 2021. Springer International Publishing. ISBN 978-3-030-86520-7.

Gatys, L. A., Ecker, A. S., and Bethge, M. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, 2016. doi: 10.1109/CVPR.2016.265.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Greydanus, S., Dzamba, M., and Yosinski, J. Hamiltonian neural networks. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/26cd8ecadce0d4efd6cc8a8725cbd1f8-Paper.pdf.

Li, X., Wong, T.-K. L., Chen, R. T. Q., and Duvenaud, D. Scalable gradients for stochastic differential equations. *International Conference on Artificial Intelligence and Statistics*, 2020.

Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=B1QRgziT-.

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., and Chen, M. Glide: Towards photorealistic image generation and editing with text-guided diffusion models, 2021. URL https://arxiv.org/abs/2112.10741.

Pascanu, R., Mikolov, T., and Bengio, Y. On the difficulty of training recurrent neural networks. *30th International Conference on Machine Learning, ICML 2013*, 11 2012.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/rezende15.html.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models, 2021. URL https://arxiv.org/abs/2112.10752.

Rosenblatt, M. Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics*, 23(3):470 – 472, 1952. doi: 10.1214/aoms/1177729394. URL https://doi.org/10.1214/aoms/1177729394.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017. doi: 10.1109/ICCV.2017.244.