# Learning Flat Latent Manifolds with VAEs

**Nutan Chen**[1], Alexej Klushyn[1], Francesco Ferroni[2], Justin Bayer[1], Patrick van der Smagt[1]

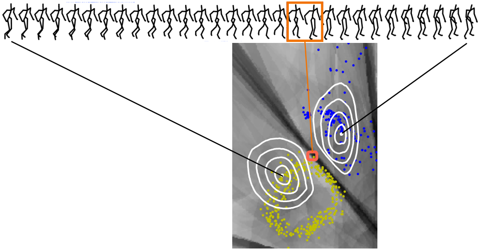
[1]Machine Learning Research Lab, Volkswagen Group, Munich, Germany

[2]Autonomous Intelligent Driving GmbH, Munich, Germany

ICML 2020

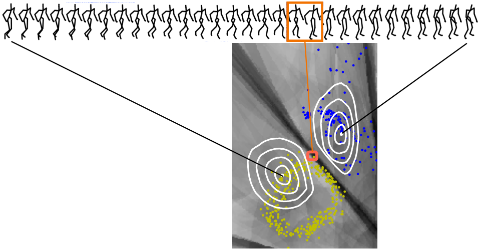
# Introduction

## Problem statement



# Introduction

## Problem statement



## The goal of this study

a latent representation, where the Euclidean metric is a proxy for the similarity between data points

# Background on Riemannian distance with VAEs

The observation-space length is defined as [CKK+18]:

$$L(\gamma) = \int_0^1 \sqrt{\dot{\gamma}(t)^T \, \mathbf{G}\big(\gamma(t)\big) \, \dot{\gamma}(t)} \, \mathrm{d}t.$$

$\gamma : [0, 1] \rightarrow \mathbb{R}^{N_\mathbf{z}}$ in the latent space
$\mathbf{G}(\mathbf{z}) = \mathbf{J}(\mathbf{z})^T \mathbf{J}(\mathbf{z})$: Riemannian metric tensor
$\mathbf{J}$: the Jacobian of the decoder
$\mathbf{z} \in \mathbb{R}^{N_z}$: latent variables
$\mathbf{x} \in \mathbb{R}^{N_x}$: observable data

# Background on Riemannian distance with VAEs

The observation-space length is defined as [CKK$^+$18]:

$$L(\gamma) = \int_0^1 \sqrt{\dot{\gamma}(t)^T \, \mathbf{G}\big(\gamma(t)\big) \, \dot{\gamma}(t)} \, \mathrm{d}t.$$

$\gamma : [0,1] \to \mathbb{R}^{N_\mathbf{z}}$ in the latent space
$\mathbf{G}(\mathbf{z}) = \mathbf{J}(\mathbf{z})^T \mathbf{J}(\mathbf{z})$: Riemannian metric tensor
$\mathbf{J}$: the Jacobian of the decoder
$\mathbf{z} \in \mathbb{R}^{N_z}$: latent variables
$\mathbf{x} \in \mathbb{R}^{N_x}$: observable data

*observation-space distance*:

$$D = \min_{\gamma} L(\gamma)$$

# Flat manifold VAEs

$$D \propto \|\mathbf{z}(1) - \mathbf{z}(0)\|_2$$

$$\mathbf{G} \propto \mathbb{1}$$

# Flat manifold VAEs

$$D \propto \|\mathbf{z}(1) - \mathbf{z}(0)\|_2$$
$$\mathbf{G} \propto \mathbb{1}$$

- flexible prior
- regularise the Jacobian of the decoder
- data augmentation in the low density area

# Flat manifold VAEs

$$\mathcal{L}_{\text{VHP-FMVAE}}(\theta, \phi, \Theta, \Phi; \lambda, \eta, c^2) = \underbrace{\mathcal{L}_{\text{VHP}}(\theta, \phi, \Theta, \Phi; \lambda)}_{\text{loss of the VHP-VAE [KCK}^{+}\text{19]}} + \eta \underbrace{\mathbb{E}_{\mathbf{x}_{i,j} \sim p_{\mathcal{D}}(\mathbf{x})} \mathbb{E}_{\mathbf{z}_{i,j} \sim q_{\phi}(\mathbf{z}|\mathbf{x}_{i,j})} \left[ \| \mathbf{G}(g(\mathbf{z}_i, \mathbf{z}_j)) - c^2 \mathbb{1} \|_2^2 \right]}_{\text{regulariser}},$$

$\eta$: hyper-parameter
$c$: scaling factor
$p_{\mathcal{D}}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}_i)$ is the empirical distribution of the data $\mathcal{D} = \{\mathbf{x_i}\}_{i=1}^{N}$

# Flat manifold VAEs

$$\mathcal{L}_{\text{VHP-FMVAE}}(\theta, \phi, \Theta, \Phi; \lambda, \eta, c^2)$$
$$= \underbrace{\mathcal{L}_{\text{VHP}}(\theta, \phi, \Theta, \Phi; \lambda)}_{\text{loss of the VHP-VAE}}$$
$$+ \ \eta \underbrace{\mathbb{E}_{\mathbf{x}_{i,j} \sim p_{\mathcal{D}}(\mathbf{x})} \mathbb{E}_{\mathbf{z}_{i,j} \sim q_{\phi}(\mathbf{z}|\mathbf{x}_{i,j})} \left[ \| \mathbf{G}(g(\mathbf{z}_i, \mathbf{z}_j)) - c^2 \mathbb{1} \|_2^2 \right]}_{\text{regulariser}},$$

scaling factor

$$c^2 = \frac{1}{N_{\mathbf{z}}} \ \mathbb{E}_{\mathbf{x}_{i,j} \sim p_{\mathcal{D}}(\mathbf{x})} \mathbb{E}_{\mathbf{z}_{i,j} \sim q_{\phi}(\mathbf{z}|\mathbf{x}_{i,j})} \left[ \mathrm{tr}(\mathbf{G}(g(\mathbf{z}_i, \mathbf{z}_j))) \right].$$

# Flat manifold VAEs

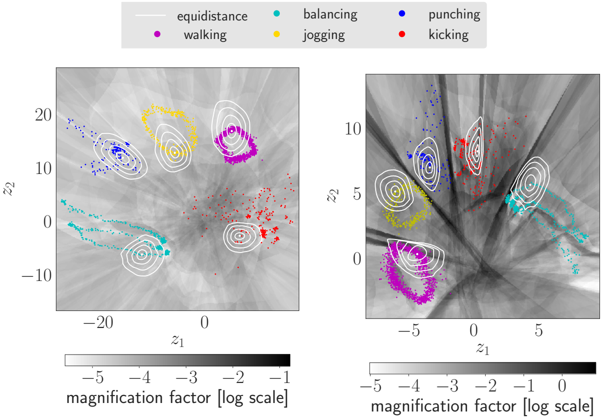
$$\begin{aligned}\mathcal{L}_{\text{VHP-FMVAE}}(\theta, \phi, \Theta, \Phi; \lambda, \eta, c^2) \\ = \underbrace{\mathcal{L}_{\text{VHP}}(\theta, \phi, \Theta, \Phi; \lambda)}_{\text{loss of the VHP-VAE}} \\ + \ \eta \underbrace{\mathbb{E}_{\mathbf{x}_{i,j} \sim p_{\mathcal{D}}(\mathbf{x})} \mathbb{E}_{\mathbf{z}_{i,j} \sim q_\phi(\mathbf{z}|\mathbf{x}_{i,j})} \left[ \| \mathbf{G}(g(\mathbf{z}_i, \mathbf{z}_j)) - c^2 \mathbb{1} \|_2^2 \right]}_{\text{regulariser}},\end{aligned}$$

mixup [ZCDLP18] in the latent space

$$g(\mathbf{z}_i, \mathbf{z}_j) = (1 - \alpha)\, \mathbf{z}_i + \alpha\, \mathbf{z}_j,$$

with $\mathbf{x}_i, \mathbf{x}_j \sim p_{\mathcal{D}}(\mathbf{x})$, $\mathbf{z}_i \sim q_\phi(\mathbf{z}|\mathbf{x}_i)$, $\mathbf{z}_j \sim q_\phi(\mathbf{z}|\mathbf{x}_j)$, and $\alpha \sim U(-\alpha_0, 1 + \alpha_0)$.

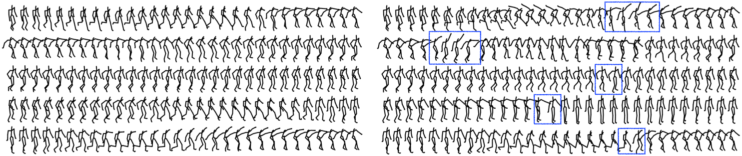
# Visualisation of equidistances on 2D latent space




(a) VHP-FMVAE (b) VHP-VAE

Round, homogeneous contour plots indicate that $\mathbf{G}(\mathbf{z}) \propto \mathbb{1}$.

# Smoothness of Euclidean interpolations in the latent space



(a) VHP-FMVAE (b) VHP-VAE

# VHP-FMVAE-SORT for MOT16 [MLTR+16] Object-Tracking Database

| Method | Type | $IDF_1\uparrow$ | IDP↑ | IDR↑ | Recall↑ | Precision↑ | FAR↓ | MT↑ |
|---|---|---|---|---|---|---|---|---|
| VHP-FMVAE-SORT $\eta = 300$ (ours) | unsupervised | 63.7 | 77.0 | 54.3 | 65.0 | 92.3 | **1.12** | 158 |
| VHP-FMVAE-SORT $\eta = 3000$ (ours) | unsupervised | 64.2 | **77.6** | 54.8 | 65.1 | **92.3** | 1.13 | 162 |
| VHP-VAE-SORT | unsupervised | 60.5 | 72.3 | 52.1 | 65.8 | 91.4 | 1.28 | 170 |
| SORT [BGO+16] | n.a. | 57.0 | 67.4 | 49.4 | 66.4 | 90.6 | 1.44 | 158 |
| DeepSORT [WBP17] | supervised | **64.7** | 76.9 | **55.8** | **66.7** | 91.9 | 1.22 | **180** |

| Method | PT↓ | ML↓ | FP↓ | FN↓ | IDs↓ | FM↓ | MOTA ↑ | MOTP ↑ | MOTAL↑ |
|---|---|---|---|---|---|---|---|---|---|
| VHP-FMVAE-SORT $\eta = 300$ (ours) | 269 | 90 | **5950** | 38592 | 616 | **1143** | 59.1 | 81.8 | 59.7 |
| VHP-FMVAE-SORT $\eta = 3000$ (ours) | 265 | 90 | 6026 | 38515 | 598 | 1163 | 59.1 | 81.8 | 59.7 |
| VHP-VAE-SORT | 266 | **81** | 6820 | 37739 | 693 | 1264 | 59.0 | 81.6 | 59.6 |
| SORT | 275 | 84 | 7643 | 37071 | 1486 | 1515 | 58.2 | **81.9** | 59.5 |
| DeepSORT | **250** | 87 | 6506 | **36747** | **585** | 1165 | **60.3** | 81.6 | **60.8** |

# VHP-FMVAE-SORT for MOT16 Object-Tracking Database

# Conclusion

- Euclidean metric is a proxy for the data similarity.
- The proposed method nears that of supervised approaches.

- Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft, *Simple online and realtime tracking*, IEEE ICIP, 2016, pp. 3464–3468.
- Nutan Chen, Alexej Klushyn, Richard Kurle, Xueyan Jiang, Justin Bayer, and Patrick van der Smagt, *Metrics for deep generative models*, AISTATS, 2018, pp. 1540–1550.
- Alexej Klushyn, Nutan Chen, Richard Kurle, Botond Cseke, and Patrick van der Smagt, *Learning hierarchical priors in VAEs*, NeurIPS (2019).
- Anton Milan, Laura Leal-Taixé, Ian Reid, Stefan Roth, and Konrad Schindler, *Mot16: A benchmark for multi-object tracking*, arXiv preprint arXiv:1603.00831 (2016).
- Nicolai Wojke, Alex Bewley, and Dietrich Paulus, *Simple online and realtime tracking with a deep association metric*, IEEE International Conference on Image Processing, 2017, pp. 3645–3649.

Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz, *mixup: Beyond empirical risk minimization*, International Conference on Learning Representations (2018).