EuroVis Workshop on Visual Analytics (2025)
H.-J. Schulz and A. Vilanova (Editors)

Evaluating Autoencoders for Parametric and Invertible Multidimensional Projections – Supplementary Material

Frederik L. Dennig[1], Nina Geyer[1], Daniela Blumberg[1], Yannick Metz[1], and Daniel A. Keim[1]

[1]University of Konstanz, Germany

1. Visual Parameter Analysis of ω, α and β

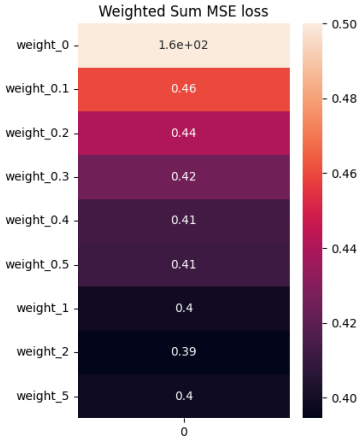

Figure 1: *Weighted sum (1:1) of MSE on reconstruction and MSE on latent space for different values of* ω *(*`weight`*) calculated with the MNIST dataset.*

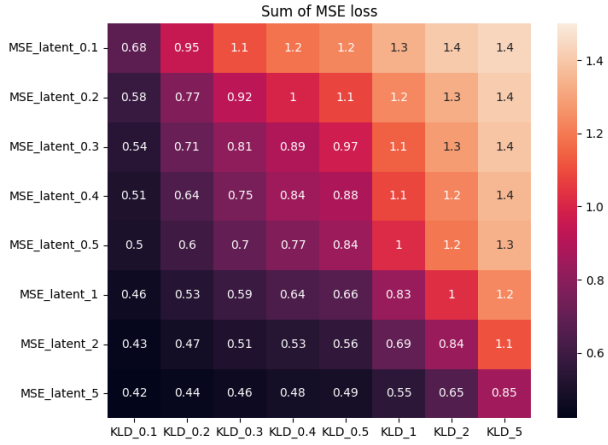

Figure 2: *Weighted sum (1:1) of MSE on reconstruction and MSE on latent space for different values of* α *(*`MSE_latent`*) and* β *(*`KLD`*) calculated with the MNIST dataset.*

Choosing ω for AEL: In our visual parameter analysis (Fig. 1) of the loss function of AEL we use the MNIST dataset. For AEL, the reconstruction loss is complemented by an additional term that encourages the latent space to align with a given projection, with a weight parameter ω controlling the balance between projection accuracy and reconstruction quality. To determine the optimal weight, we conducted a parameter scan by evaluating the reconstruction error and latent space conformity across various values. Our experiments revealed that setting the weight to 0.5 consistently achieved low reconstruction error on test data, making it an effective choice. Notably, lower weight values led to reduced error for inverse projections but increased error for parametric projections, highlighting the trade-off in selection. We recommend choosing a weight within the range of 0.4 to 5.0, as these values generally result in similarly low errors while maintaining a stable latent space representation.

Choosing α and β for VAEL: In our visual parameter analysis of VAEL, inspired by Higgins et al. [HMP*17], using the MNIST dataset, we explored different weight combinations to balance reconstruction quality and latent space conformity (Fig. 2). The loss function includes two weight parameters: one controlling the inverse projection smoothness and another influencing the overall weighted sum of reconstruction errors. Through extensive sampling, we found that setting the first parameter α to 1.0 and the second parameter β to 0.1 achieved generally good performance on test data. Notably, a smaller first parameter leads to a better balance between parametric and inverse projection errors, while increasing the second parameter raises the weighted sum of errors. However, the first parameter α can be increased up to 5.0 without negatively impacting reconstruction quality. Since the first parameter directly affects the smoothness of the inverse projection, we recommend evaluating its value on a case-by-case basis to optimize performance.

2. Converge of Training

We analyze the convergence of neural network training for each dataset by visualizing its loss curve, which plots the loss value against training epochs. We train each model 10 times with different initializations to avoid outliers. Generally, a well-structured loss curve typically shows a steady decline, indicating effective learning and parameter optimization. However, rapid fluctuations or plateaus may signal issues such as an improper learning rate, insufficient training time, or poor initialization. By examining the shape and trends of the loss curve, we can fine-tune hyperparameters, adjust the learning rate, or apply regularization techniques to ensure smooth convergence and optimal model performance.

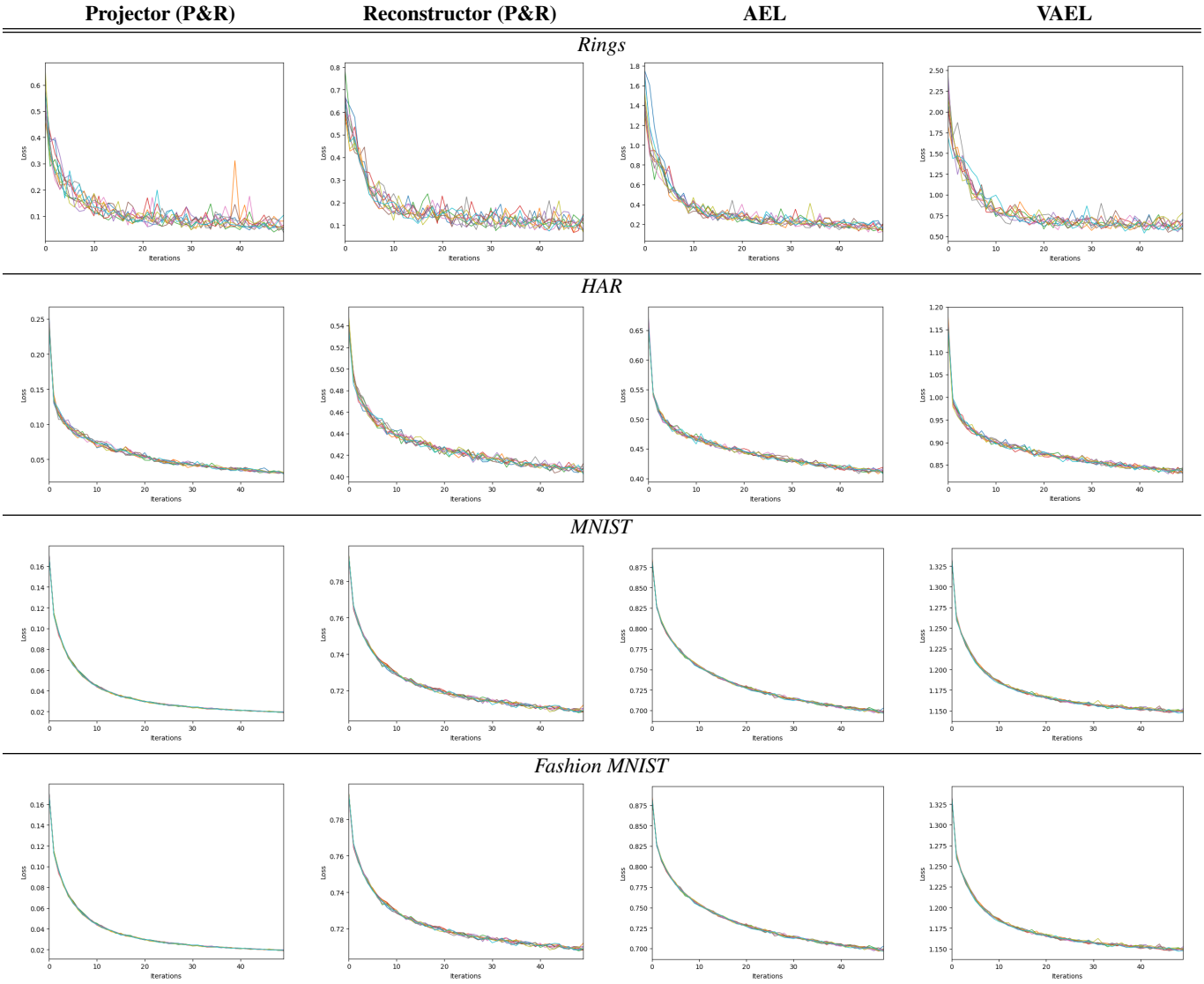

Figure 3: *Neural network training convergence for each dataset. We train each model 10 times to avoid outliers.*

3. Ground-Truth Reconstruction

When comparing the original images of *MNIST* and *Fashion MNIST* to their reconstructions through inverse projections, noticeable artifacts emerge that affect visual clarity and detail. In *MNIST*, which consists of simple handwritten digits, the reconstructed images generally retain their structure but may exhibit slight blurriness around the edges, particularly in curved regions. This fuzziness results from the model's attempt to generalize across similar digit shapes. In contrast, *Fashion MNIST* images, which feature more complex textures and varying silhouettes of clothing items, tend to show greater ambiguity in their reconstructions. Fine details, such as patterns or sharp edges, may appear smoothed out or distorted, making it harder to distinguish between similar categories like shirts and dresses. These artifacts highlight the challenge of maintaining high reconstruction fidelity, especially when dealing with intricate and diverse features present in *Fashion MNIST* compared to the simpler structures of *MNIST*.

3.1. MNIST

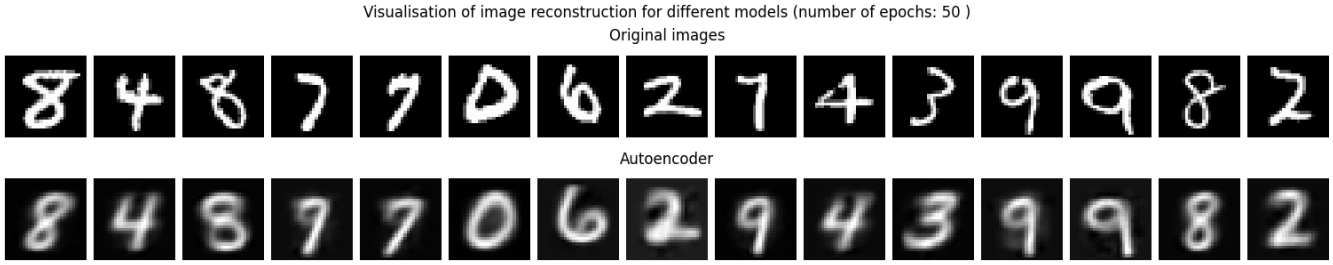

Figure 4: *Reconstruction of samples in the MNIST dataset after projection with the P&R.*

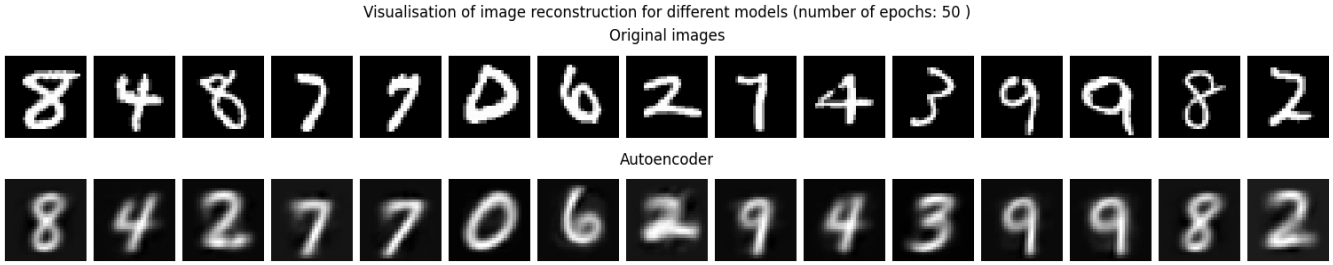

Figure 5: *Reconstruction of samples in the MNIST dataset after projection with the AEL.*

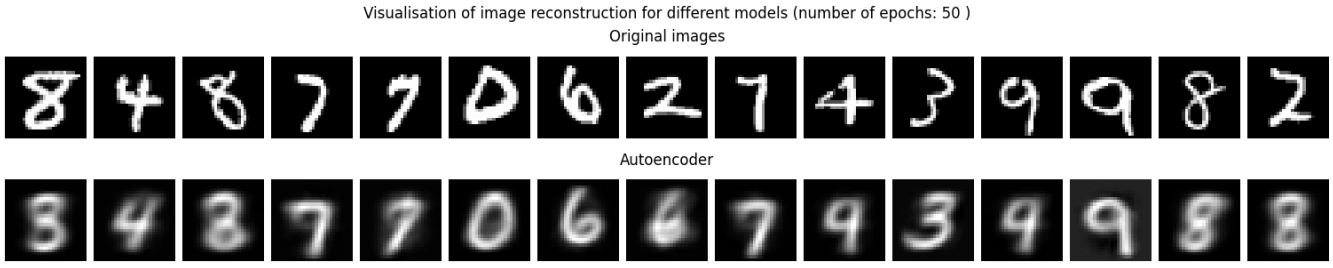

Figure 6: *Reconstruction of samples in the MNIST dataset after projection with the VAEL.*

3.2. Fashion MNIST

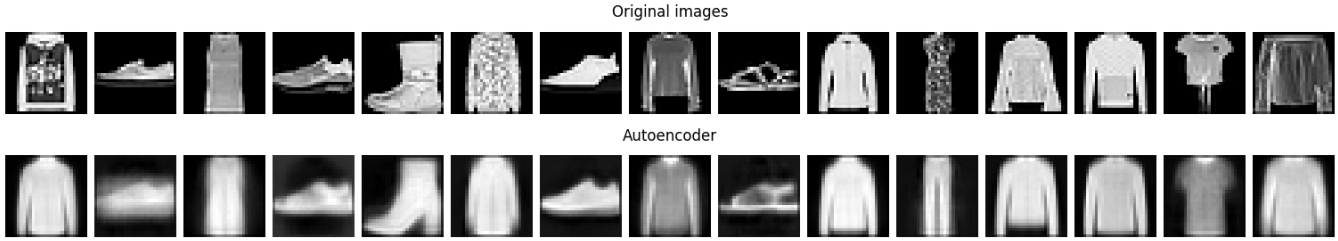

Figure 7: *Reconstruction of samples in the Fashion MNIST dataset after projection with the P&R.*

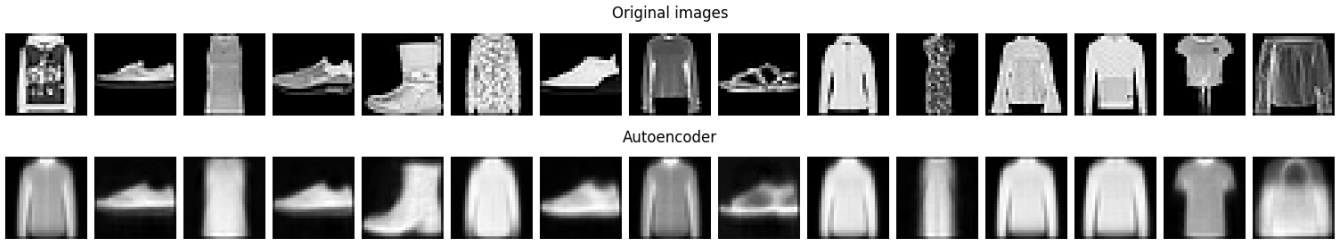

Figure 8: *Reconstruction of samples in the Fashion MNIST dataset after projection with the AEL.*

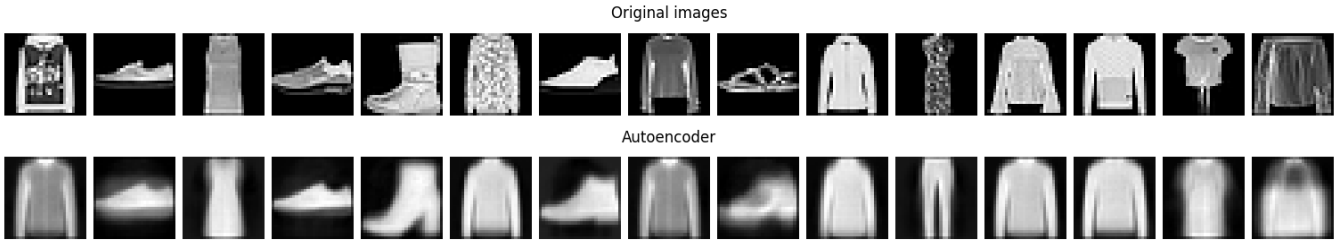

Figure 9: *Reconstruction of samples in the Fashion MNIST dataset after projection with the VAEL.*

4. Decision Maps

We use *decision maps* [EAS*21] to partition the projected space into regions of one distinct classification. This method employs a classifier (e.g., logistic regression) trained on test samples to classify the reconstructed samples. The resulting classification is visualized as distinct zones in a 2D map, allowing us to visually assess the differences in the stability of inverse projections.

Fig. 10 shows decision maps of the inverse projection. For the P&R applied on the *Rings* dataset, classification decision zones do not align with the actual data classes of the projected samples. In contrast, *HAR* and *MNIST* projections are well captured by the classifications, producing coherent and connected mappings. *Fashion MNIST* classification mappings are less coherent, reflecting inconsistencies in class mapping to the latent space. The decision maps of the AEL show better mappings for the rings dataset. With respect to *HAR* and *MNIST*, the classification is less consistent with the projection of the test sample but still coherent and connected. For the *Fashion MNIST* dataset, there is more ambiguity and an artifact in the lower left corner, indicating that the reconstruction might lead to unexpected results. The VAEL leads to accurate results for the rings dataset, but for the other datasets, it creates less precise edges between the classes and more misclassifications. Altogether, for the higher-dimensional datasets, the P&R gives the most precise results, whereas for the lower-dimensional dataset, the VAEL works best.

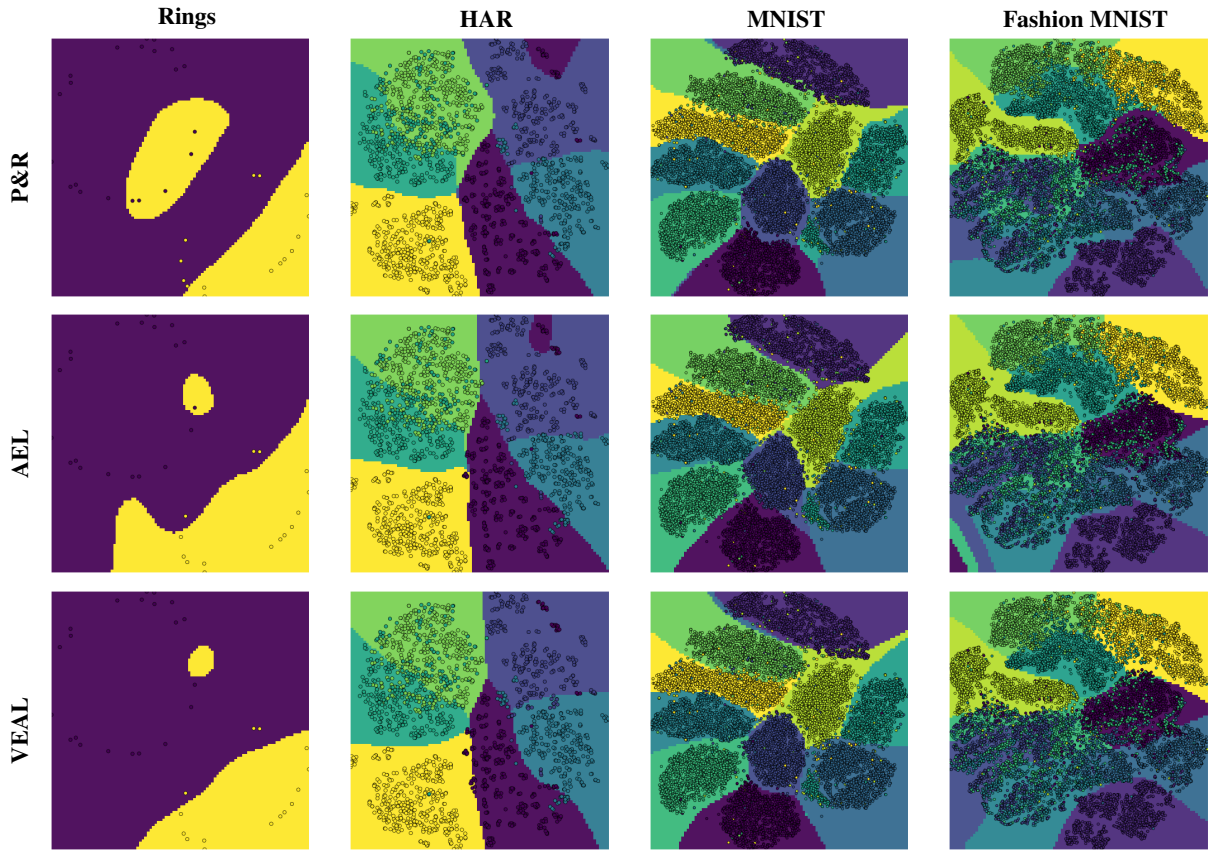

Figure 10: *Decision maps of the three inverse projection methods for four datasets. The regions show the result of a classifier, i.e., logistic regression, classifying inverse projected points of the projection space.*

References

[EAS*21] ESPADOTO, MATEUS, APPLEBY, GABRIEL, SUH, ASHLEY, et al. "UnProjection: Leveraging Inverse-Projections for Visual Analytics of High-Dimensional Data". *IEEE Trans. Vis. Comput. Graph.* 29.2 (2021), 1559–1572. DOI: 10.1109/TVCG.2021.3125576 5.

[HMP*17] HIGGINS, IRINA, MATTHEY, LOIC, PAL, ARKA, et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". *5th Int. Conf. Learn. Represent.* 2017 1.