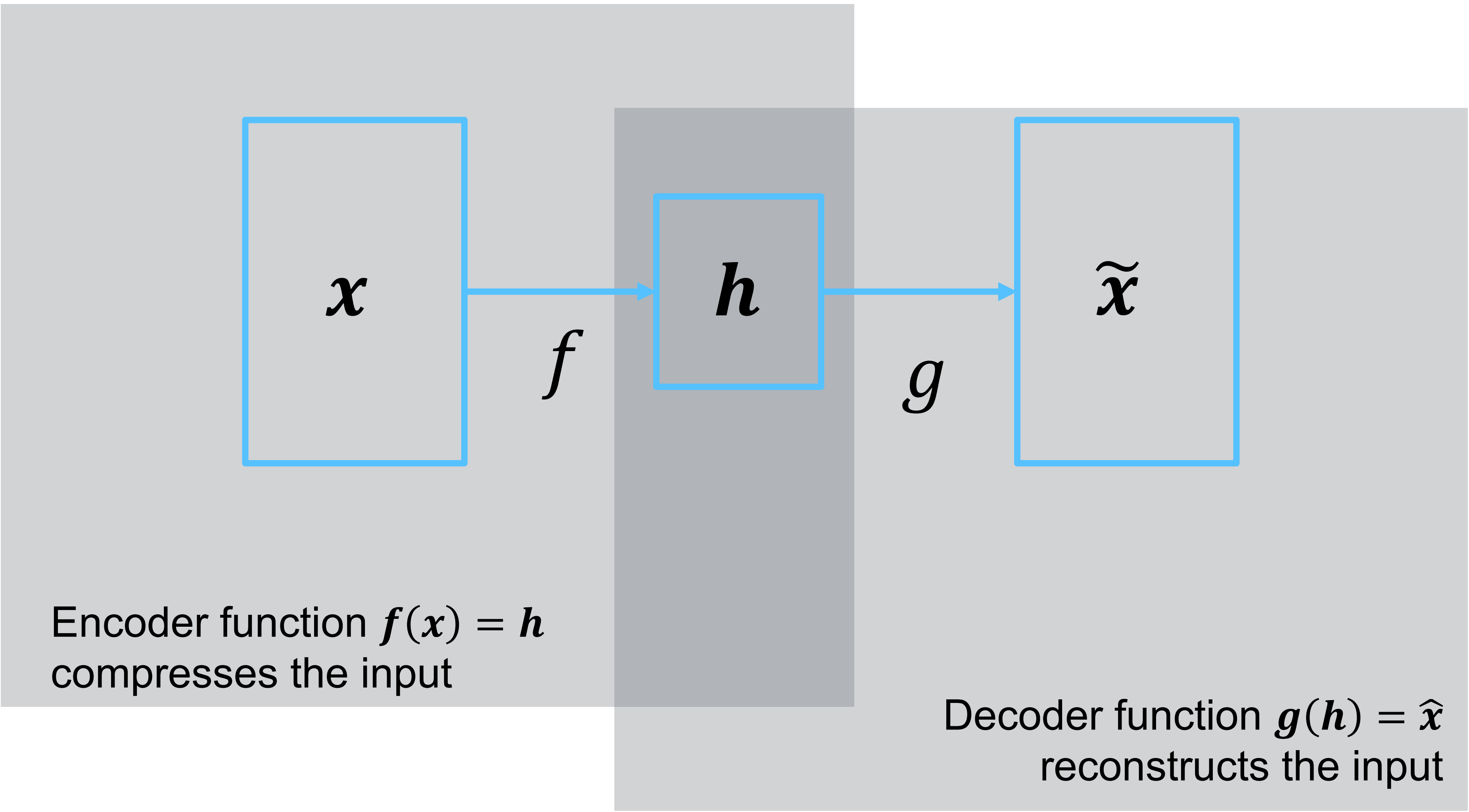
Autoencoders and their **applications** in computational biology

Marianna Rapsomaniki
Computational Systems Biology group
IBM Research Zurich
contact: aap@zurich.ibm.com

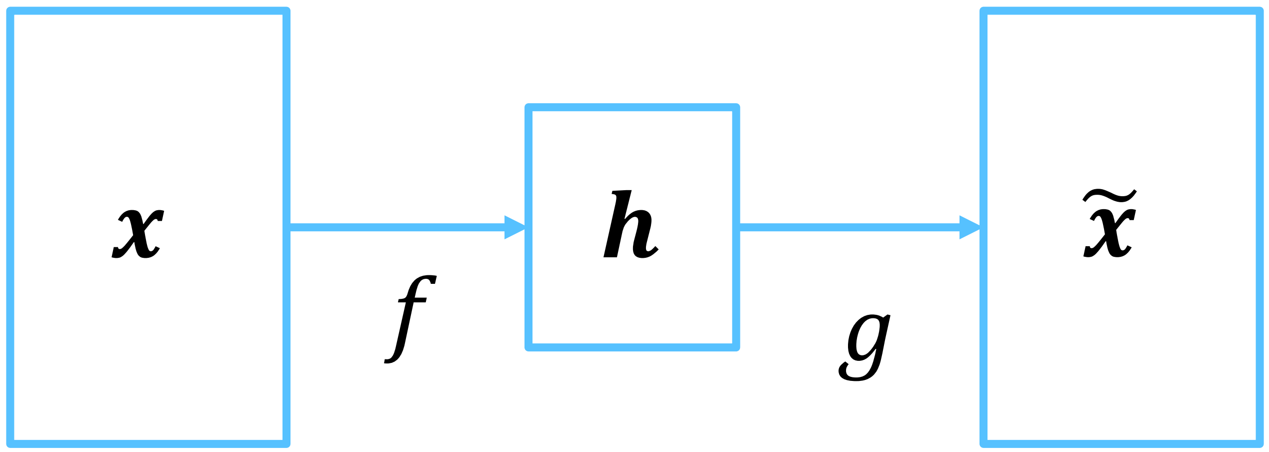
Overview

- Intro to autoencoders
- Main variations - examples
- Applications in single-cell analysis
- Hands-on exercise

Autoencoders

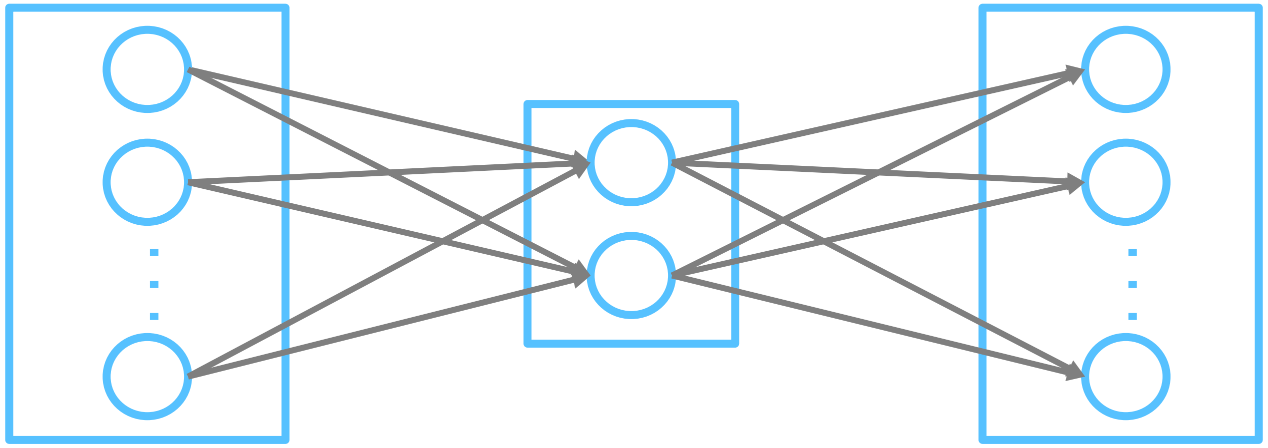


Autoencoders



- Non-linear activation functions → complex relationships
- Lossy reconstruction → most salient features of the data

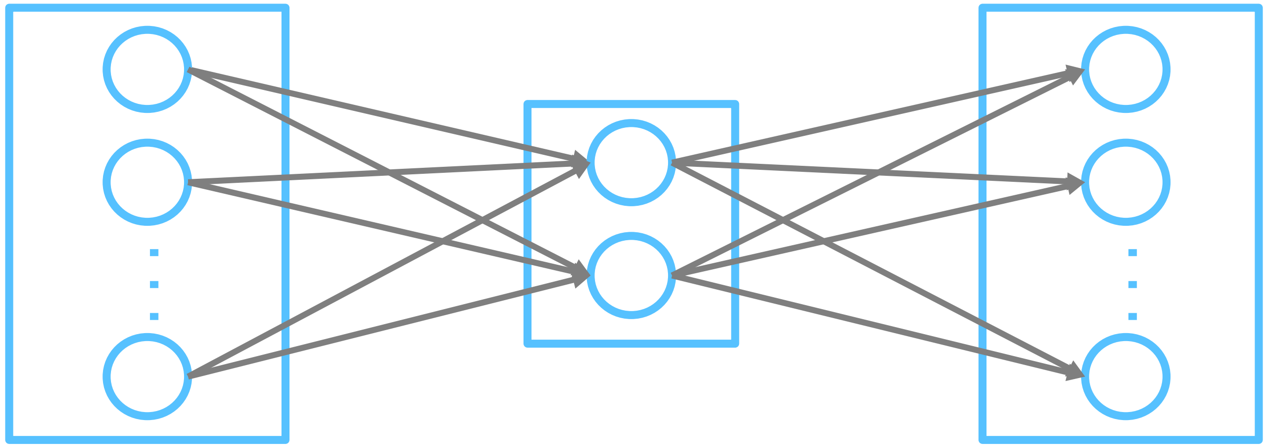
Autoencoders



Trained using **backpropagation**

$$mse = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - g\big(f(x_i)\big)\right)^2$$

Autoencoders

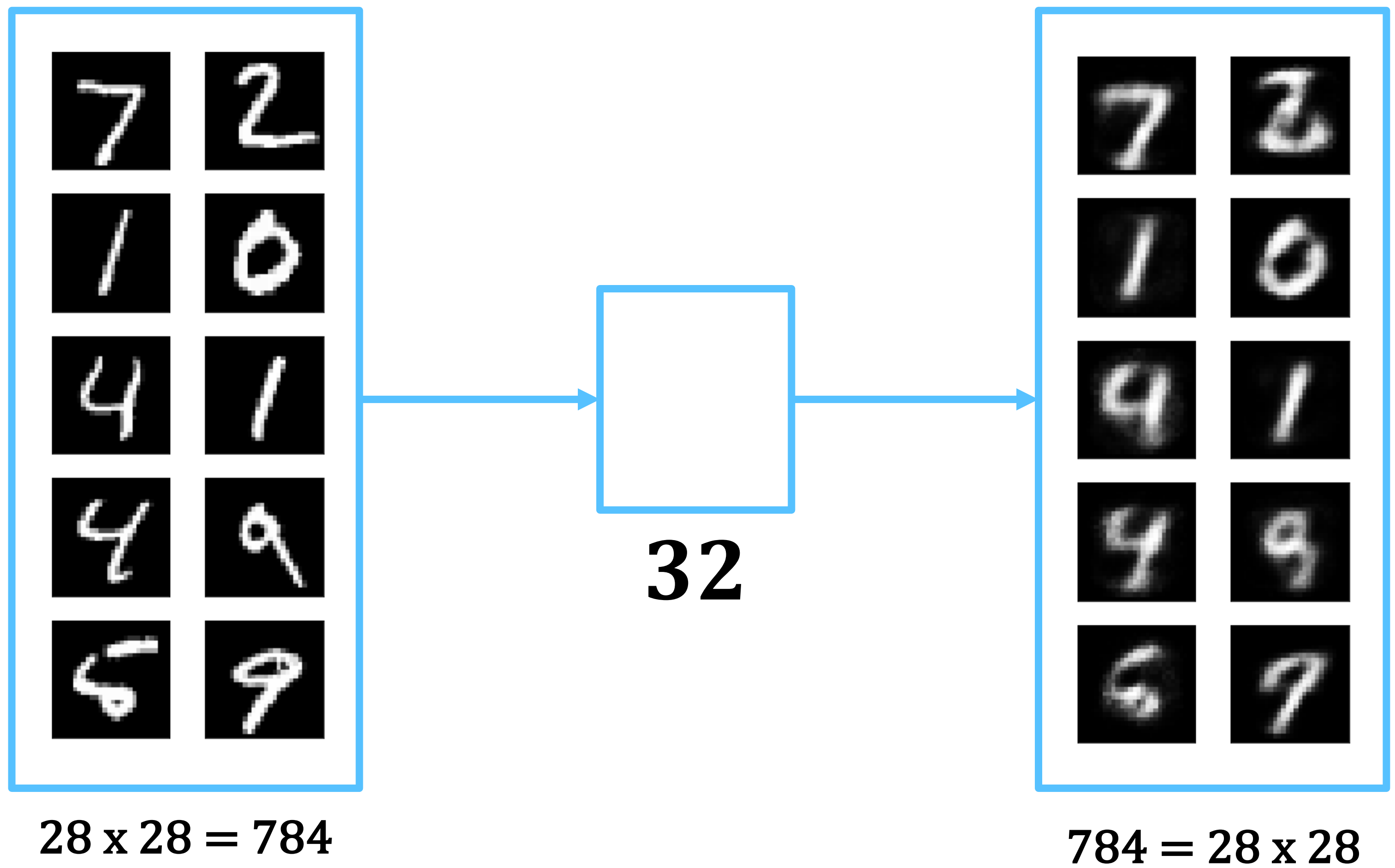


→ Data compression

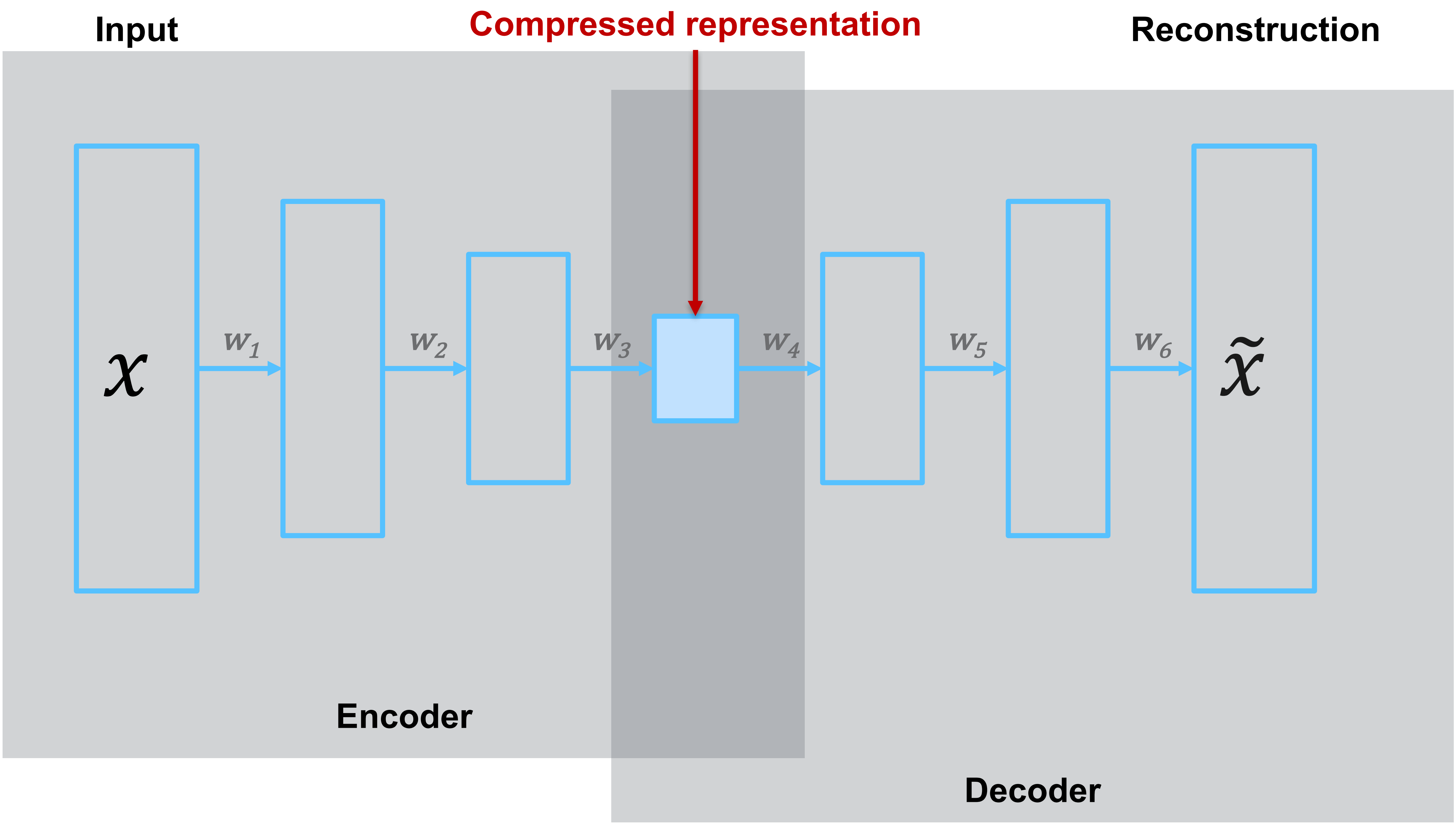
→ Experiment reconstruction

→ Dimensionality reduction

Example 1: simple AE



Deep autoencoders

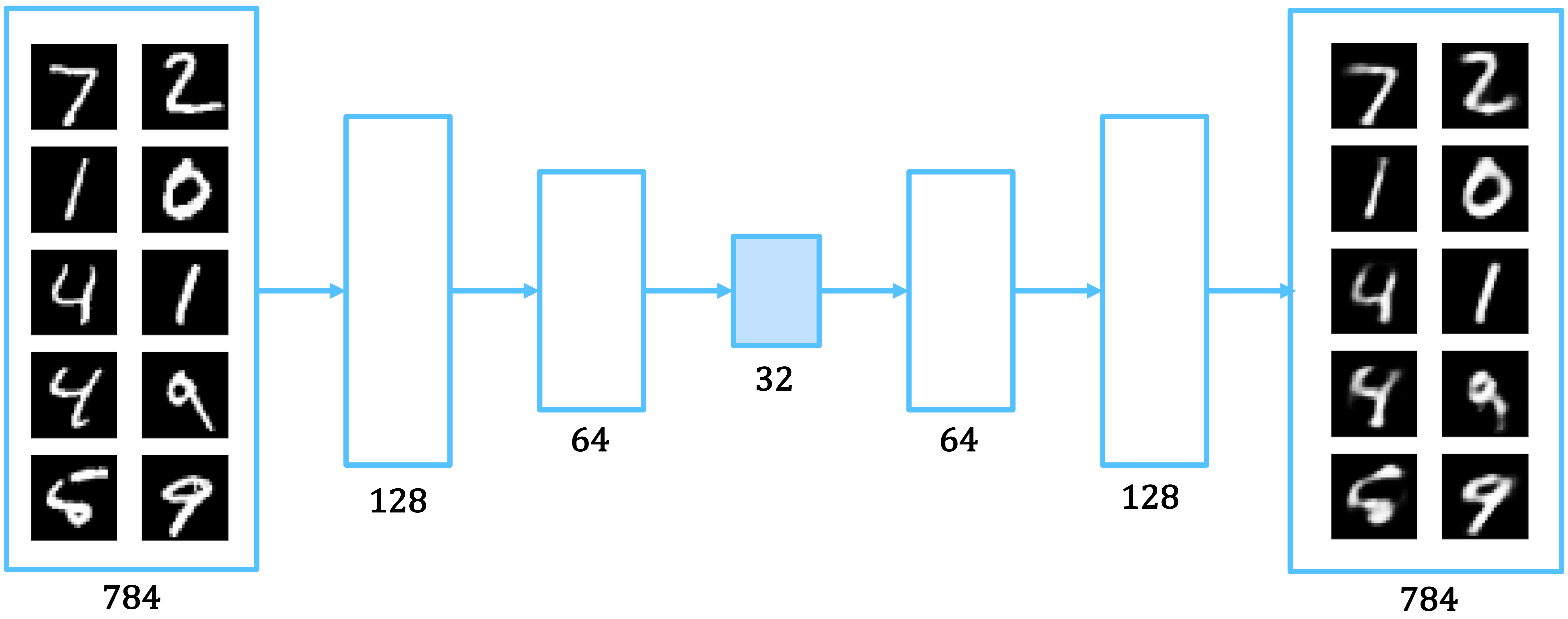


Hierarchy of representation

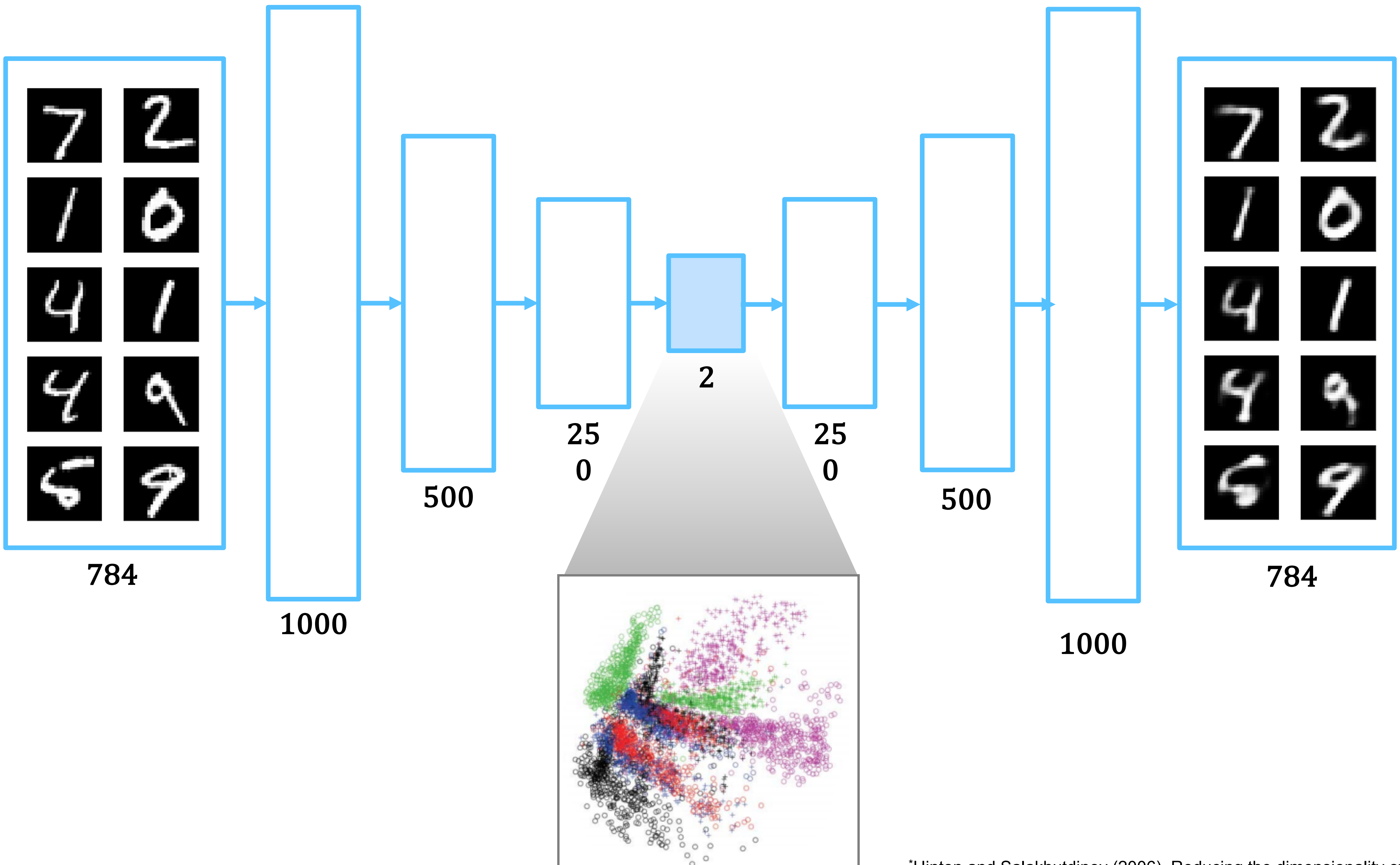
Stacking – pretraining*

*Hinton, G. E. and Salakhutdinov, R. R. (2006), Reducing the dimensionality of data with neural networks. Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006

Example 2: deep AE



Example 3: deep AE + dimensionality reduction

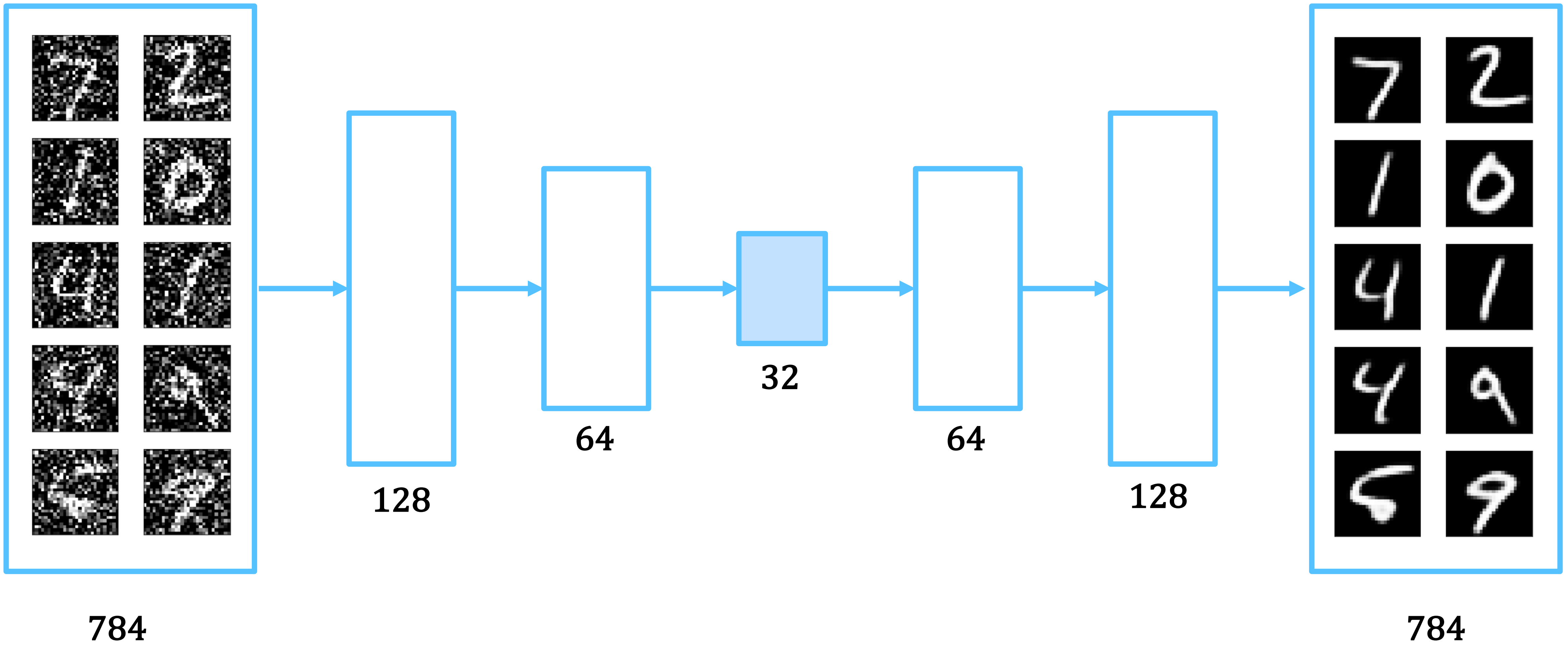


*Hinton and Salakhutdinov (2006), Reducing the dimensionality of data with neural networks. Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006

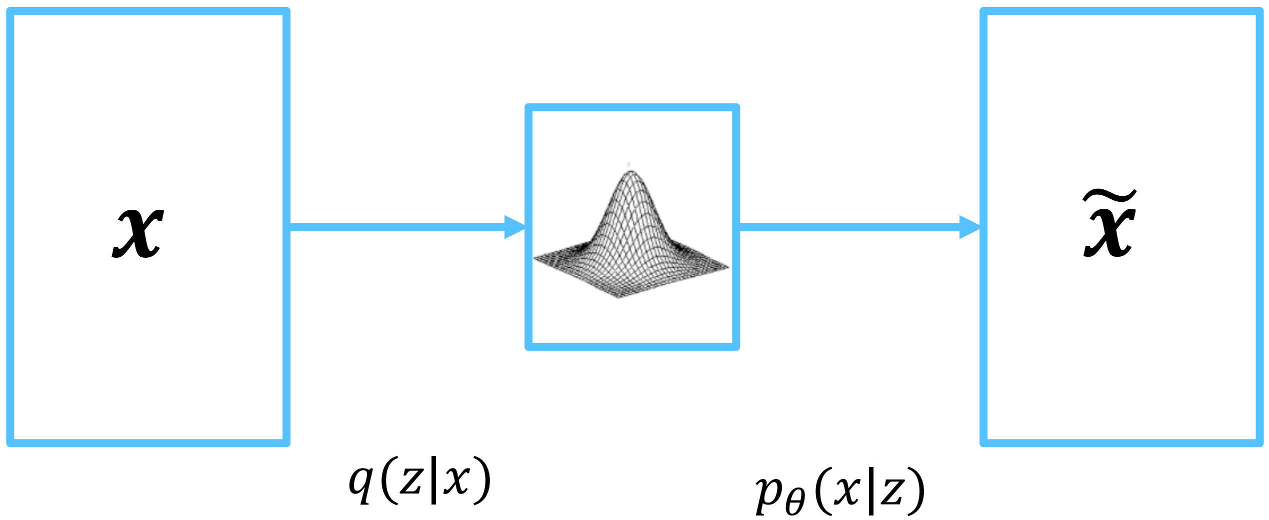
Different flavors

- **Undercomplete**: the “vanilla” flavor
- **Convolutional:** convolutions as hidden layers
- **Sparse**: impose sparsity constraint on hidden units
- **Denoising**: corrupt input with noise to increase robustness
- **Variational**: generative models, probabilistic spin

Example 4: denoising AE



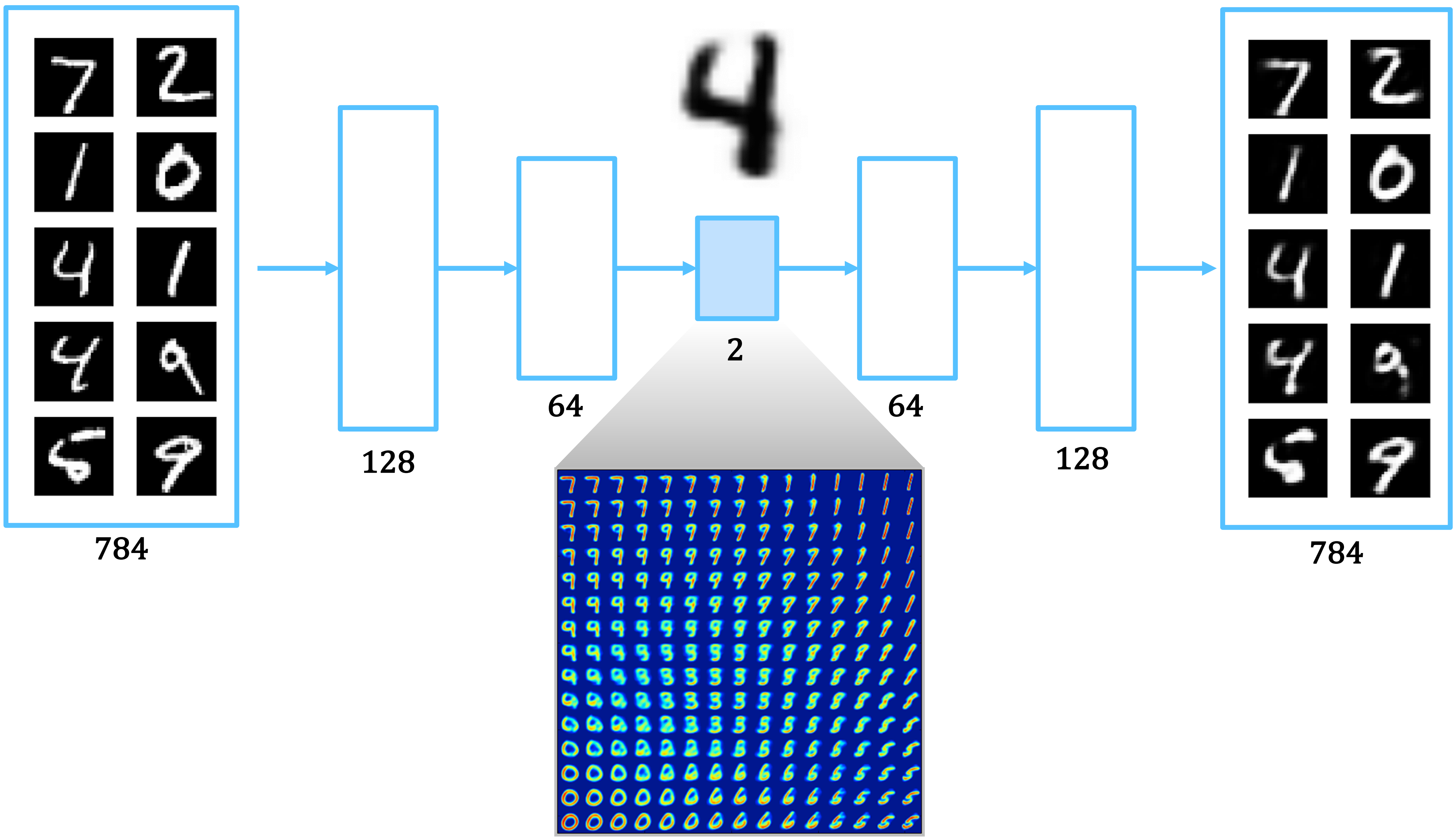
Variational Autoencoders



Encoder: maps input to posterior distributions over latent space

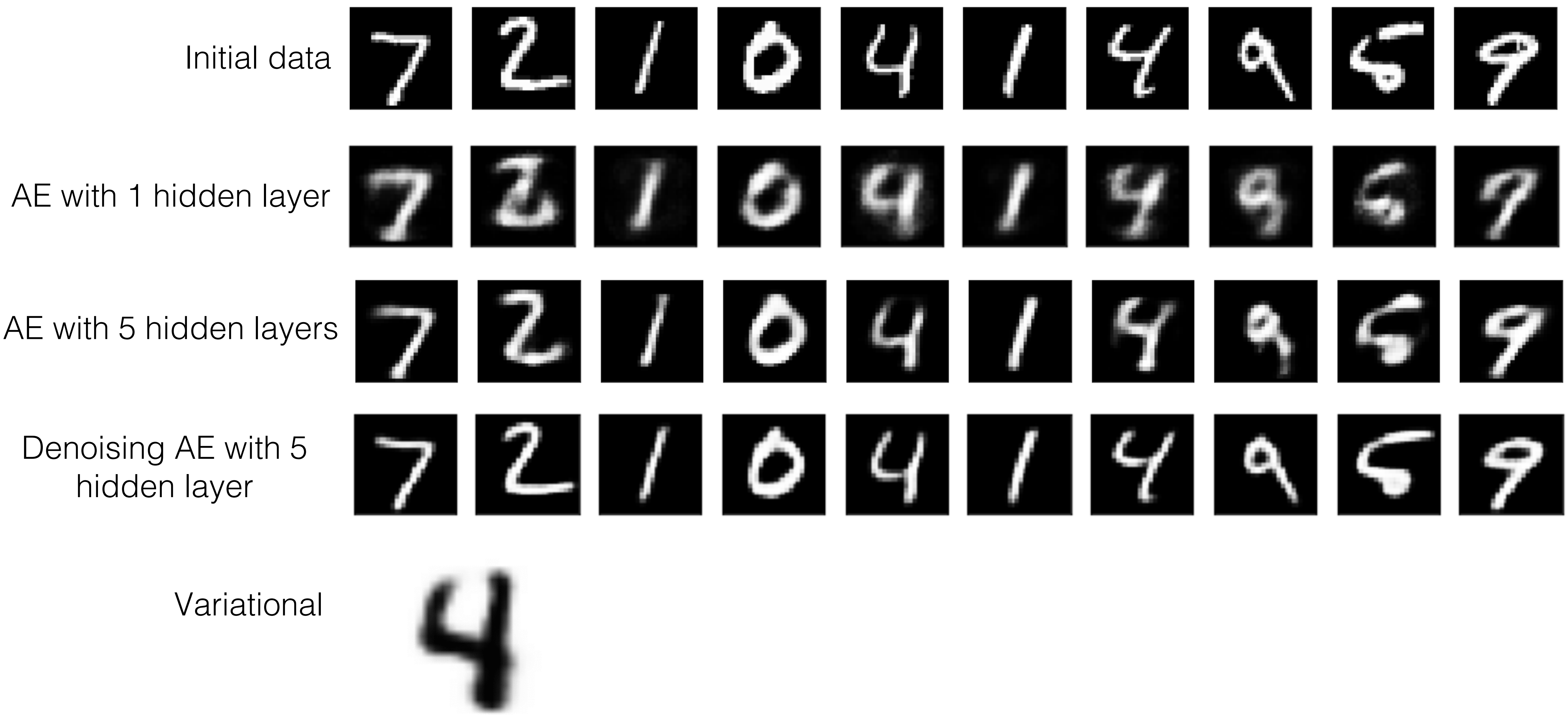
Decoder: maps coordinates back to distributions over the original space

Example 5: variational AE



Sources: http://blog.fastforwardlabs.com/2016/08/22/under-the-hood-of-the-variational-autoencoder-in.html
https://blog.keras.io/building-autoencoders-in-keras.html

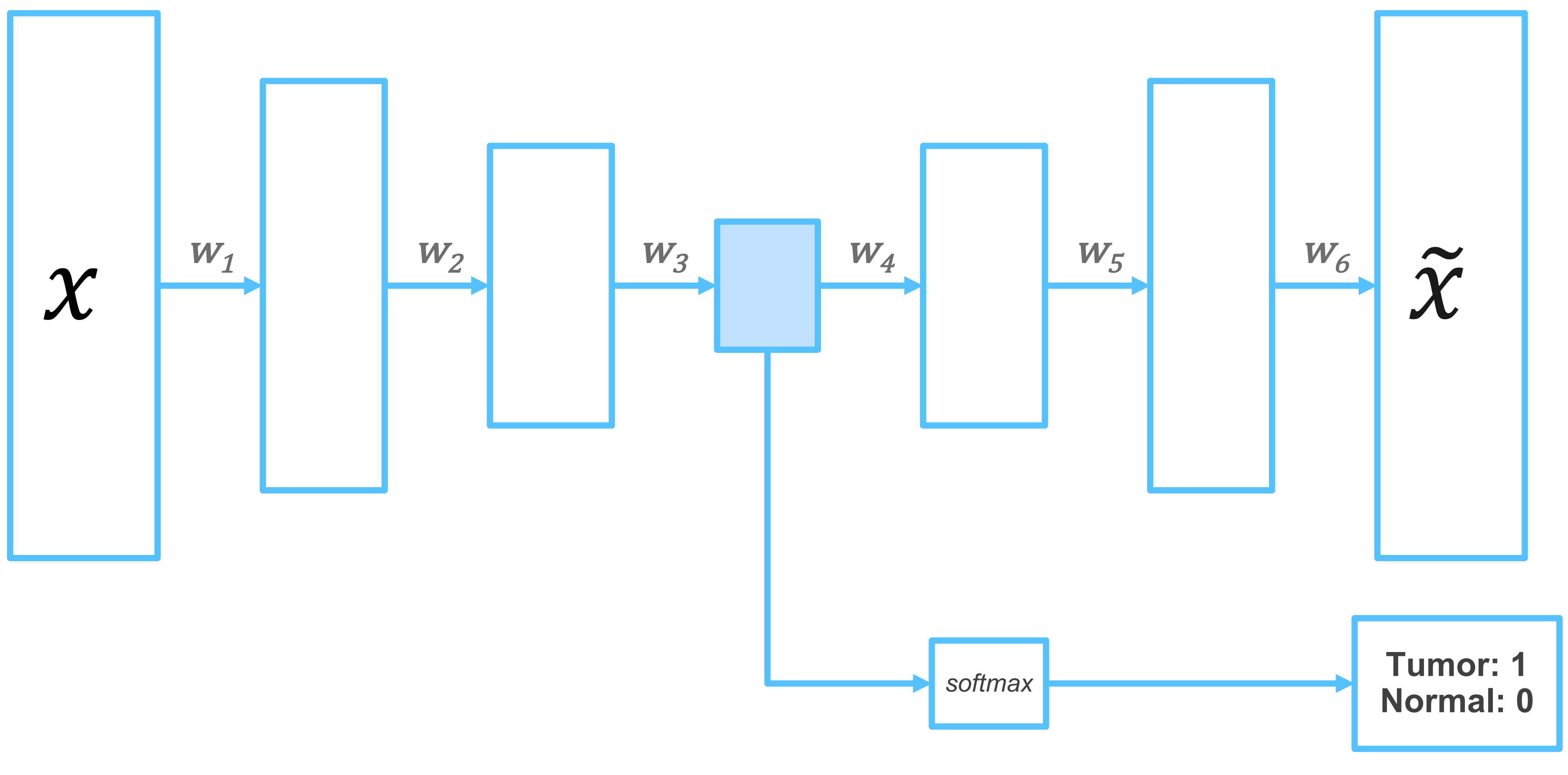
Comparison



Initial data
AE with 1 hidden layer
AE with 5 hidden layers
Denoising AE with 5 hidden layer
Variational

Extensions

Modified network architectures allow for hybrid unsupervised and supervised learning



Autoencoders for automatic gating

DeepCyTOF: Automated Cell Classification of Mass Cytometry Data by Deep Learning and Domain Adaptation

Huamin Li [1,*], Uri Shaham [2,*], Yi Yao [3], Ruth Montgomery [3], and Yuval Kluger [1,4,5†]

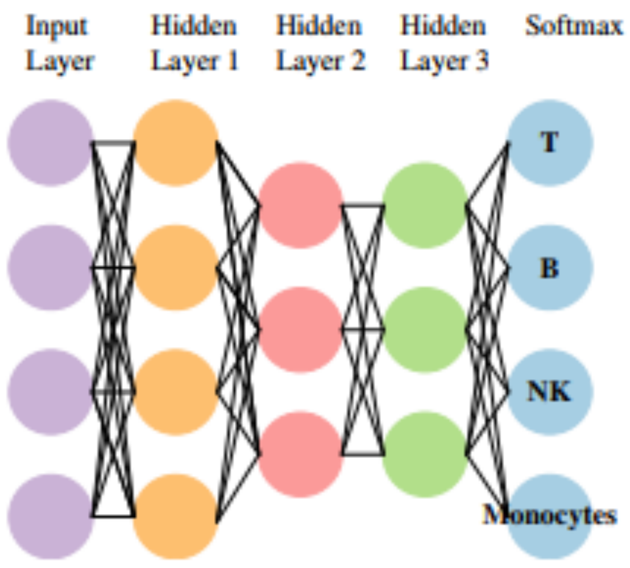


Figure 1: Stacked autoencoder for gating cell populations

Table 2: Summary of results for the cell identification challenge. The numbers in parentheses represent 95% confidence intervals

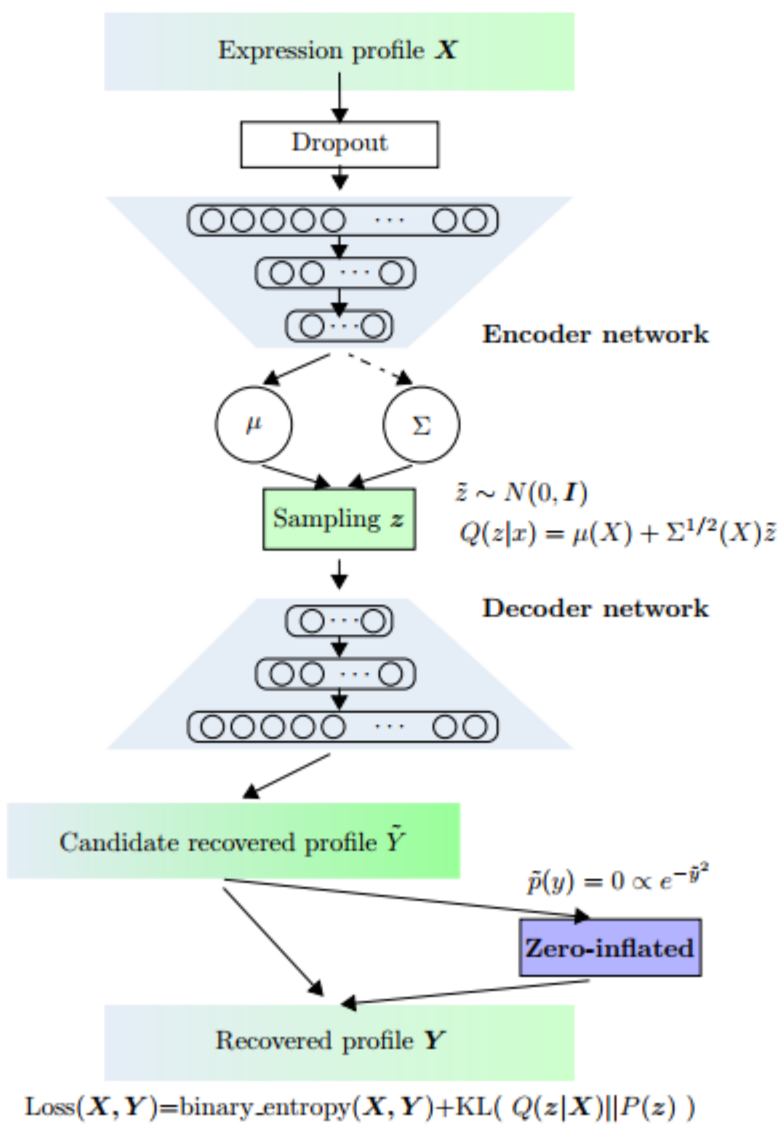
$F = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$	Stacked autoencoder	Competition's winner
GvHD[a]	0.98 (0.97,0.99)	0.92 (0.88,0.95)
DLBCL	0.97 (0.95, 0.99)	0.95 (0.93, 0.97)
HSCT	0.98 (0.96, 0.99)	0.98 (0.96, 0.99)
WNV	0.98 (0.97, 0.99)	0.96 (0.94, 0.97)
ND	0.98 (0.96, 0.99)	0.94 (0.92, 0.95)

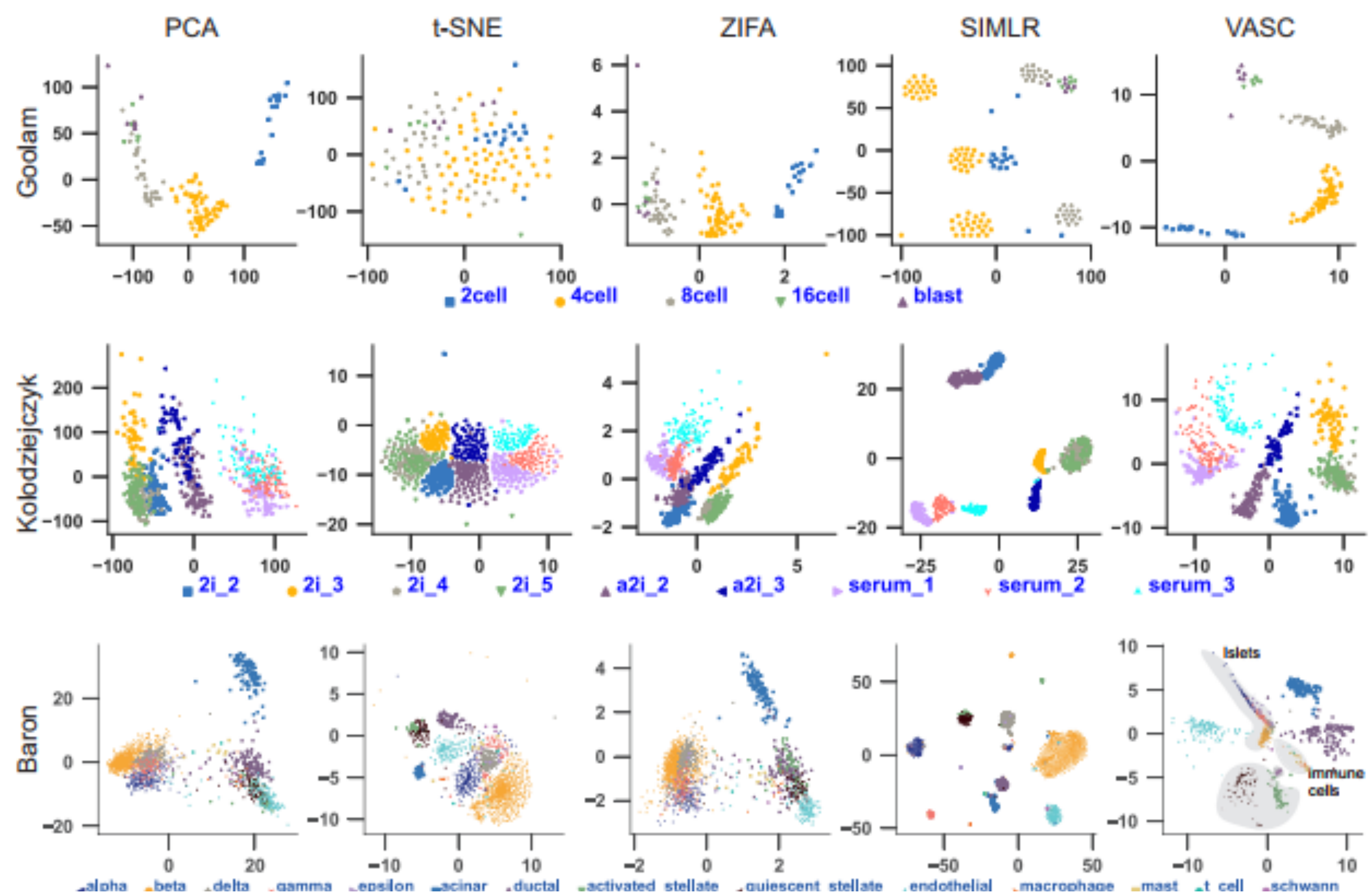
[a] Datasets: graft-versus-hist disease (GvHD); diffuse large B-cell lymphoma (DLBCL); symptomatic West Nile virus (WNV); normal donors (ND); hematopoietic stem cell transplant (HSCT).

Autoencoders for unsupervised learning

VASC: dimension reduction and visualization of single cell RNA sequencing data by deep variational autoencoder

Dongfang Wang[1], Jin Gu[1,#]





Hands-on exercise

- Characterizing cell populations from single-cell mas cytometry data

- Open your jupyter notebook **ae.ipynb**

Learning resources

Papers and blogs:

- Review: http://msb.embopress.org/content/12/7/878
- Awesome list: https://github.com/terryum/awesome-deep-learning-papers
- Variational Bayes: https://arxiv.org/pdf/1312.6114.pdf
- Sparse autoencoders: https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf
- Colah's blog: http://colah.github.io

Tutorials:

- https://blog.keras.io/building-autoencoders-in-keras.html

Courses:

- http://cs231n.github.io/neural-networks-1/
- https://stats385.github.io/readings
- http://slazebni.cs.illinois.edu/spring17/#resources
- https://www.udacity.com/course/deep-learning-nanodegree-foundation--nd101