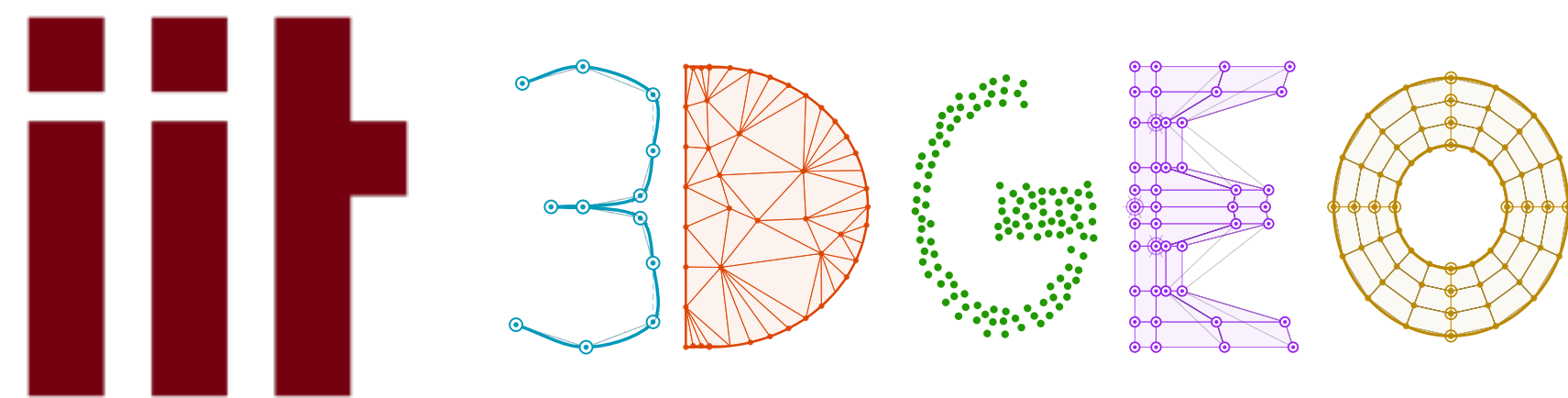
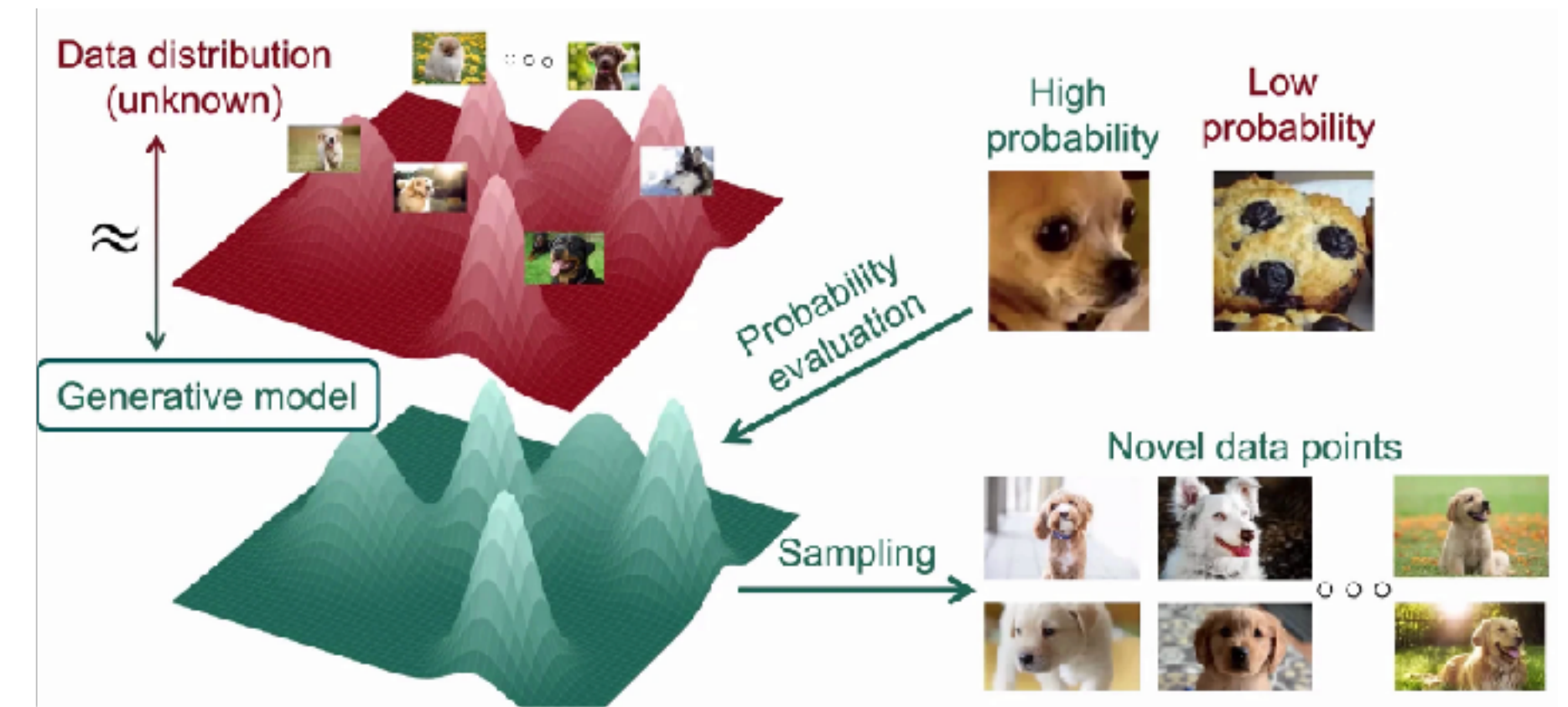


6. Előadás: Generatív Modellezés Alapjai

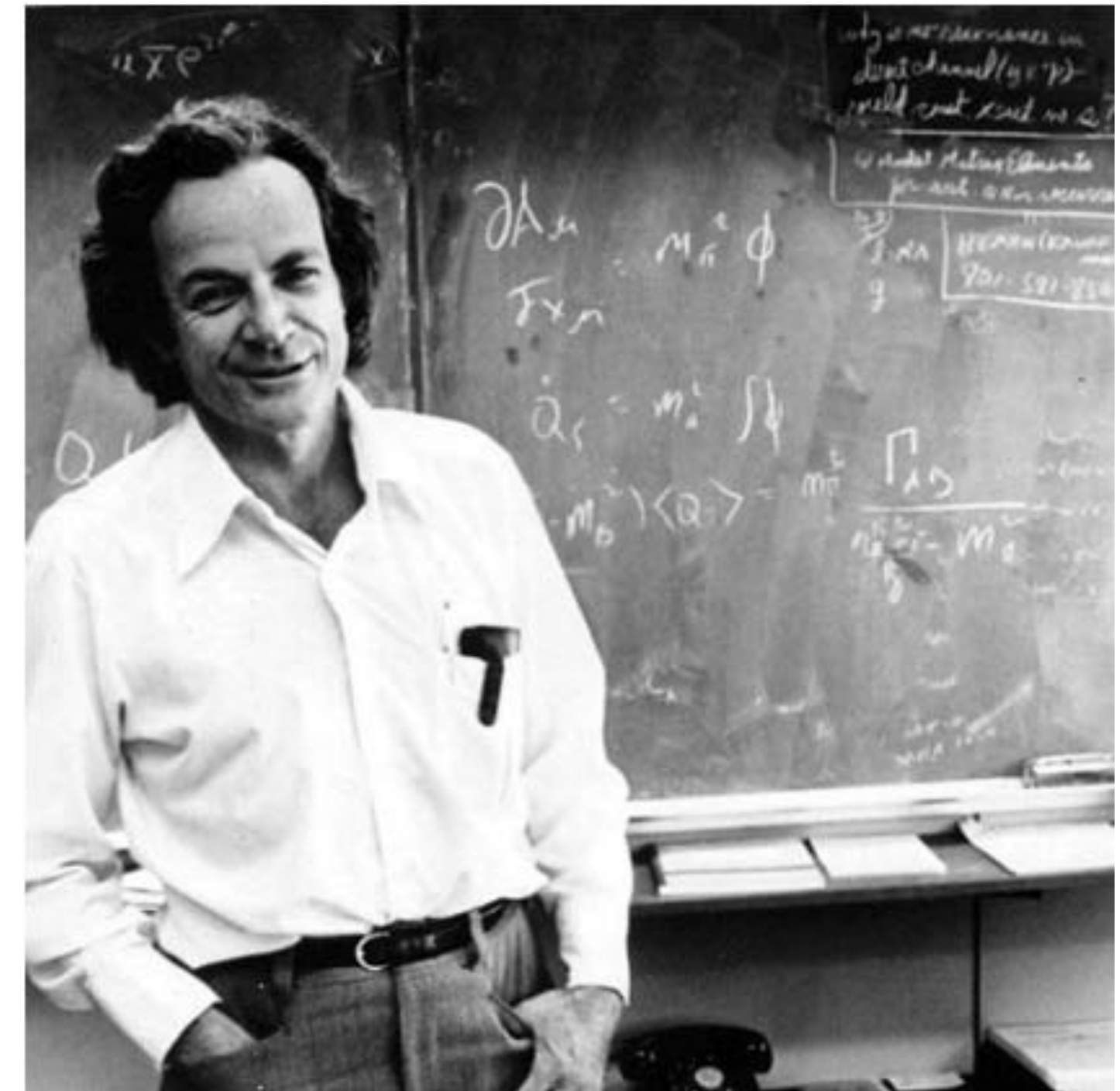
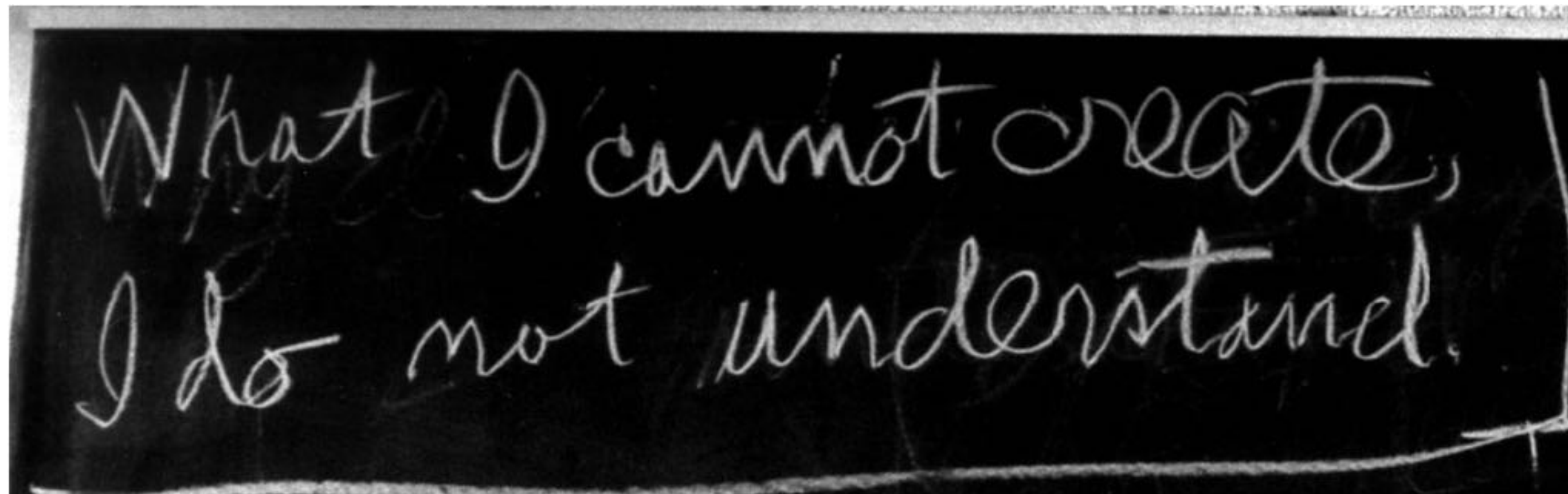
Generatív AI és Inverz Módszerek a Képszintézisben
BME-VIK IIT, 2026



Dr. Vaitkus Márton

Generatív Modellezés

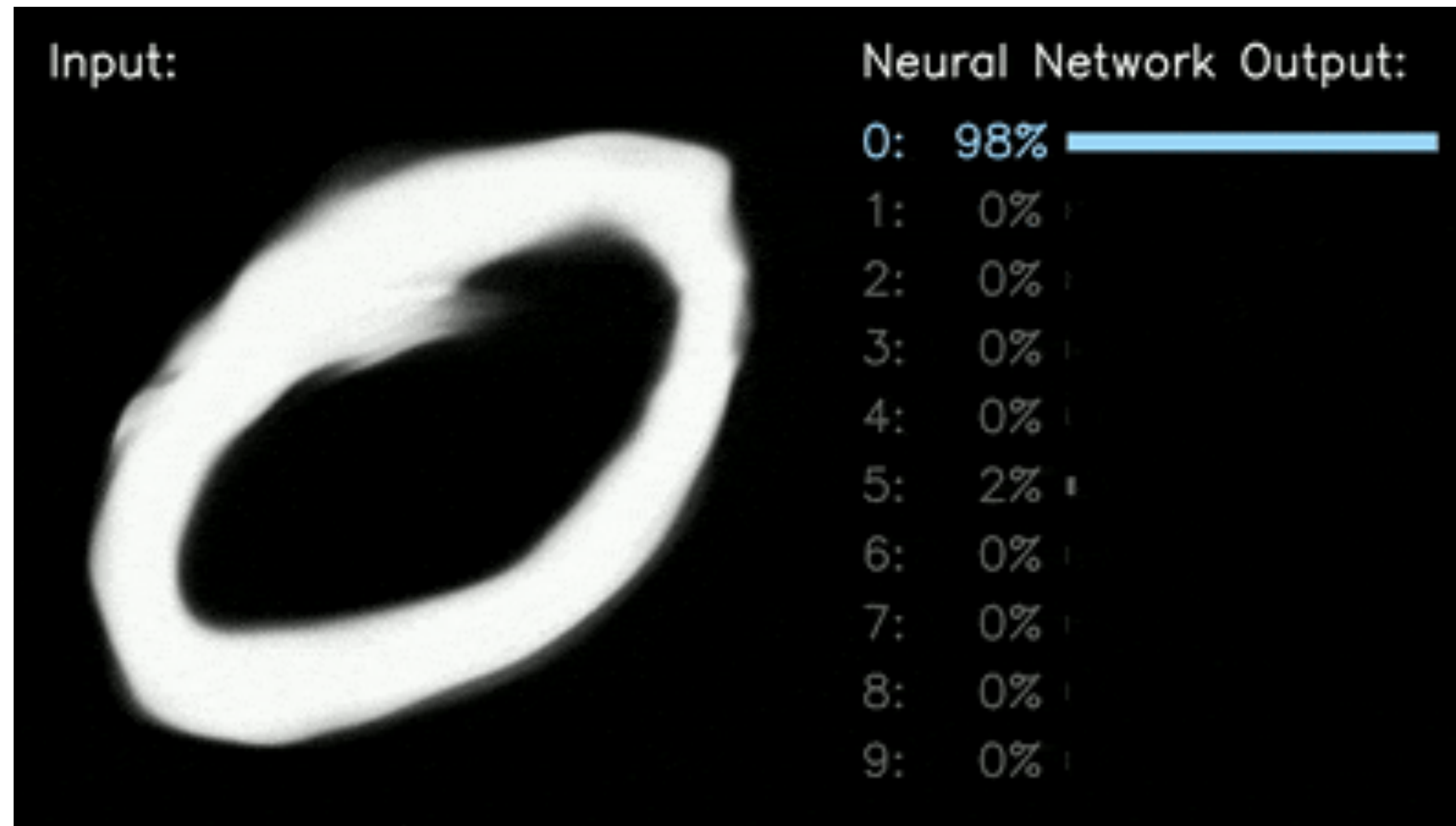
Motiváció



Richard P. Feynman (1918-1988)

Generatív Modellezés

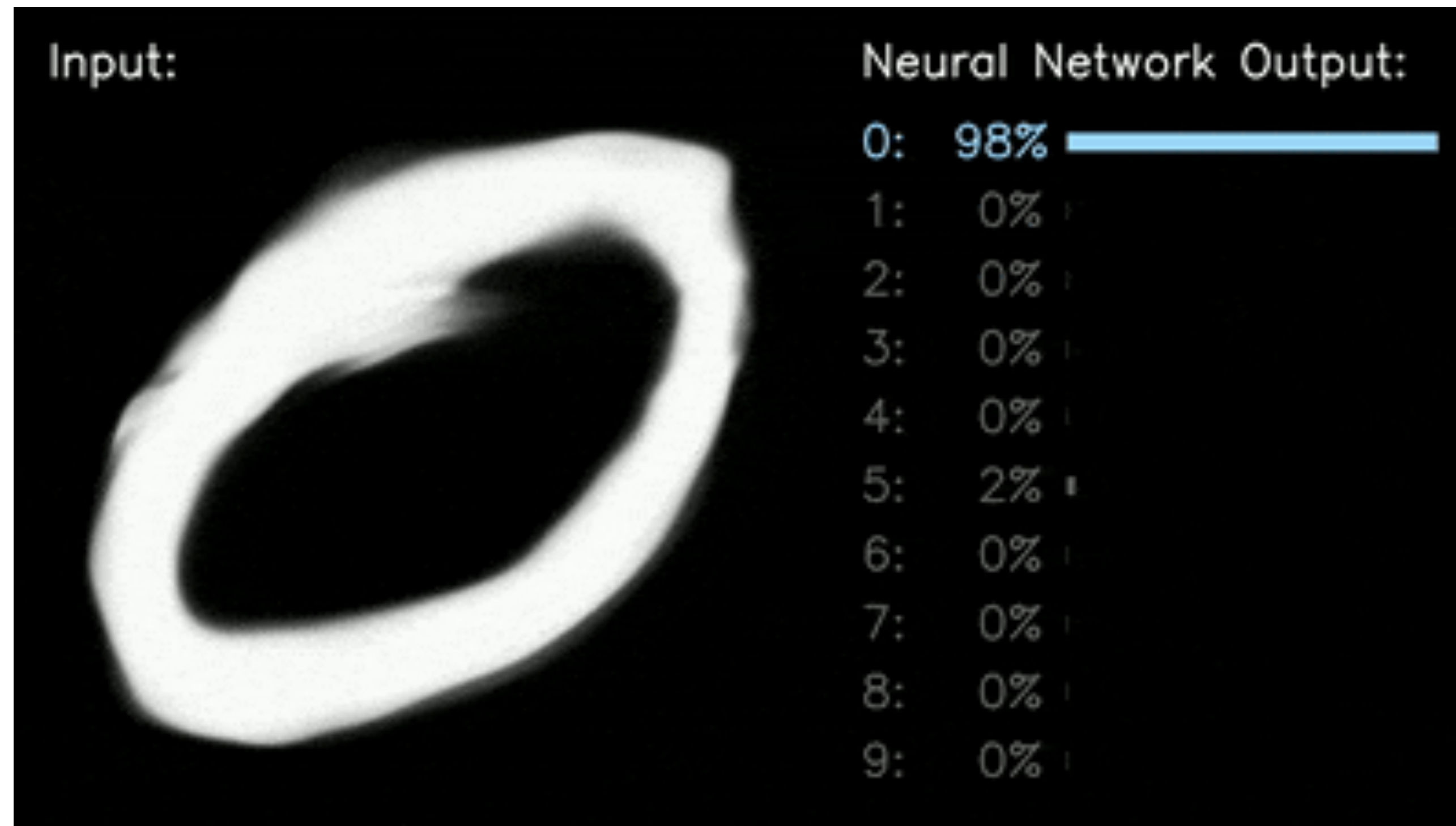
Motiváció



Amit eddig csináltunk: diszkriminatív modellezés

Generatív Modellezés

Motiváció



Amit eddig csináltunk: diszkriminatív modellezés

Generatív Modellezés

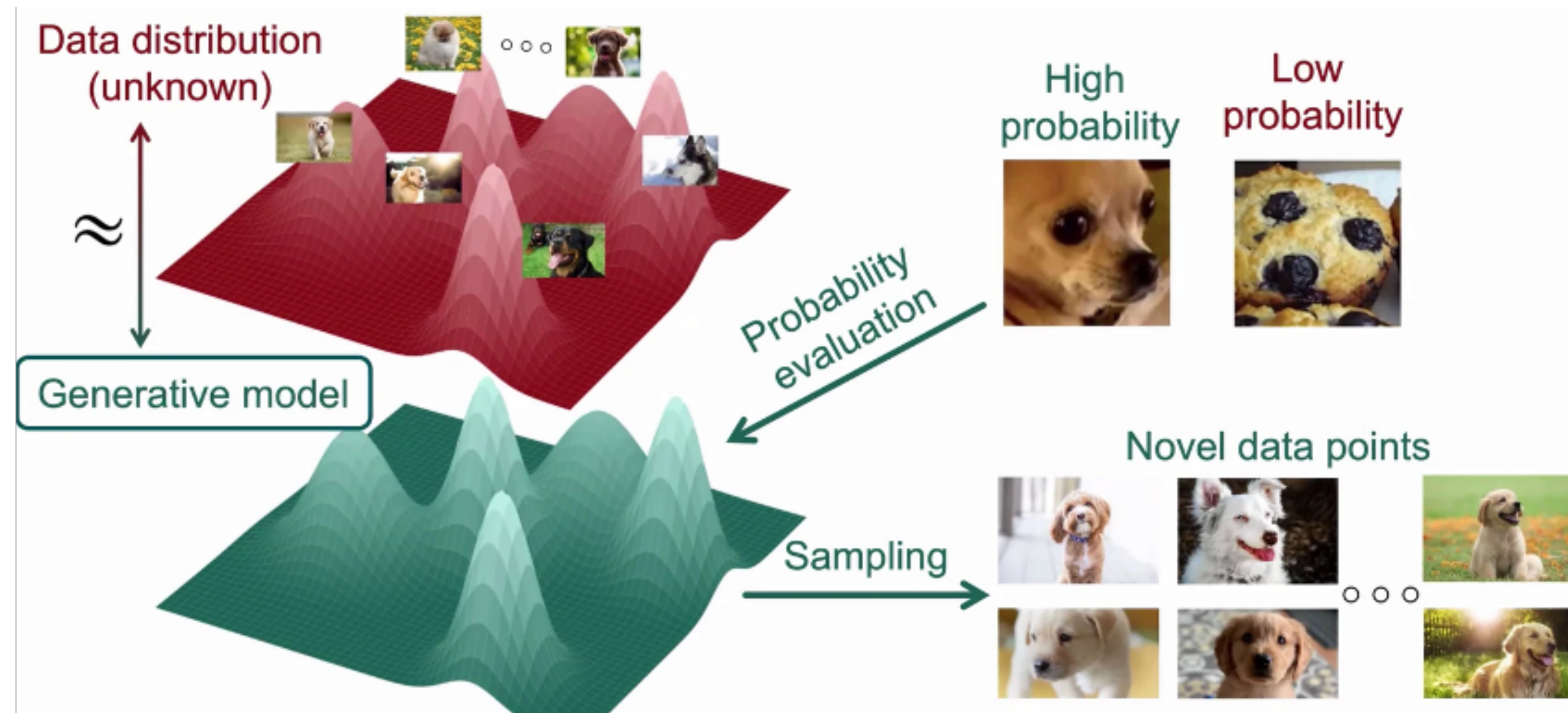
Motiváció



Mostantól új célt tűzünk ki: generatív modellezés

Generatív Modellezés

Praktikus Definíció



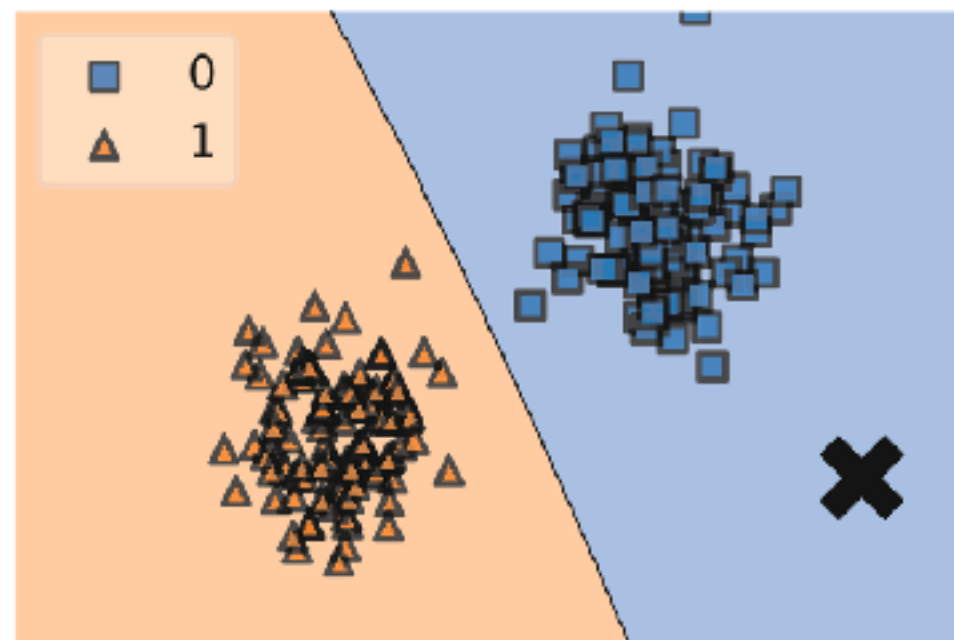
Generatív modell – cél: egy adathalmaz eloszlását modellezni és új mintákat generálni!

Generatív Modellezés

Formális Definíció

Diszkriminatív modell:

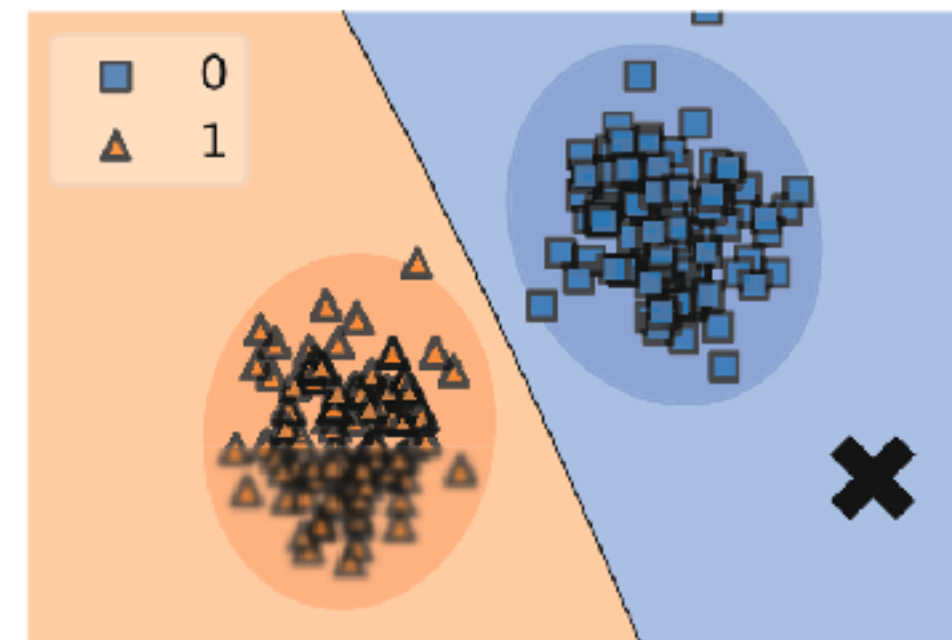
$$p(\text{címke} \mid \text{kép})$$



felügyelt tanítás!
(címkézett adatok kellene!)

Generatív modell:

$$p(\text{címke, kép}) = p(\text{címke} \mid \text{kép}) \cdot p(\text{kép})$$



Generatív modell
(*felügyelet nélküli* tanítás!):

$$p(\text{kép})$$

- “Nyers” adatokon lehet tanulni!
- Új mintákat lehet generálni!

Generatív Modellezés

Explicit vs. Implicit modellek

- **Explicit** generatív modell: explicit módon reprezentálja a $p(x)$ eloszlás sűrűségfüggvényét, amiből mintavételezéssel generál új mintákat
 - Példák: VAE, auto-regresszió, diffúzió
- **Implicit** generatív modell: csak $x \sim p(x)$ mintákat tud generálni, az eloszlás sűrűségfüggvényét nem (vagy csak implicit módon) modellezi
 - Példák: GAN

Generatív Modellelés

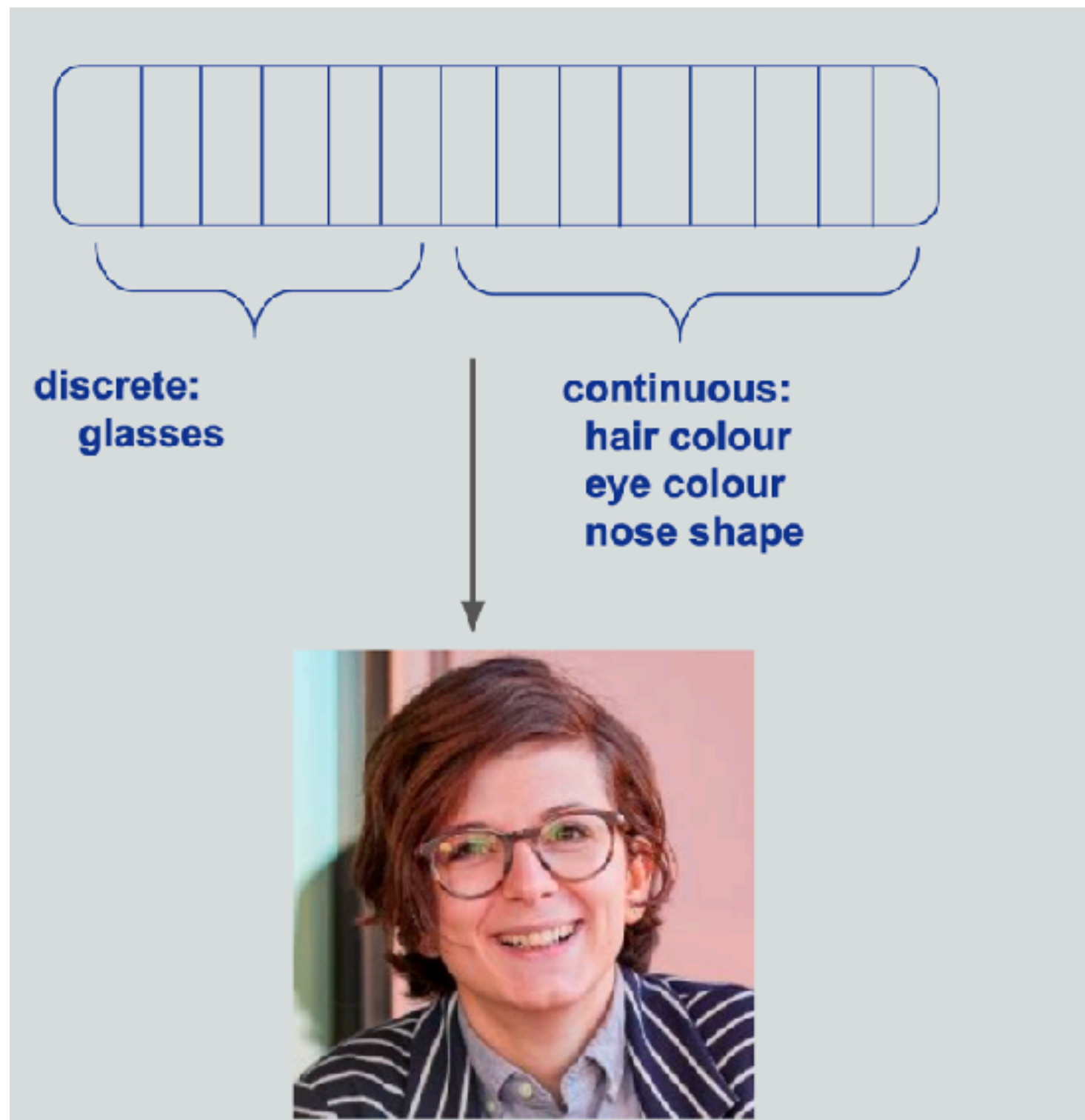
Látens változók

Generálás

látens változók



adatok



Inferencia



adatok

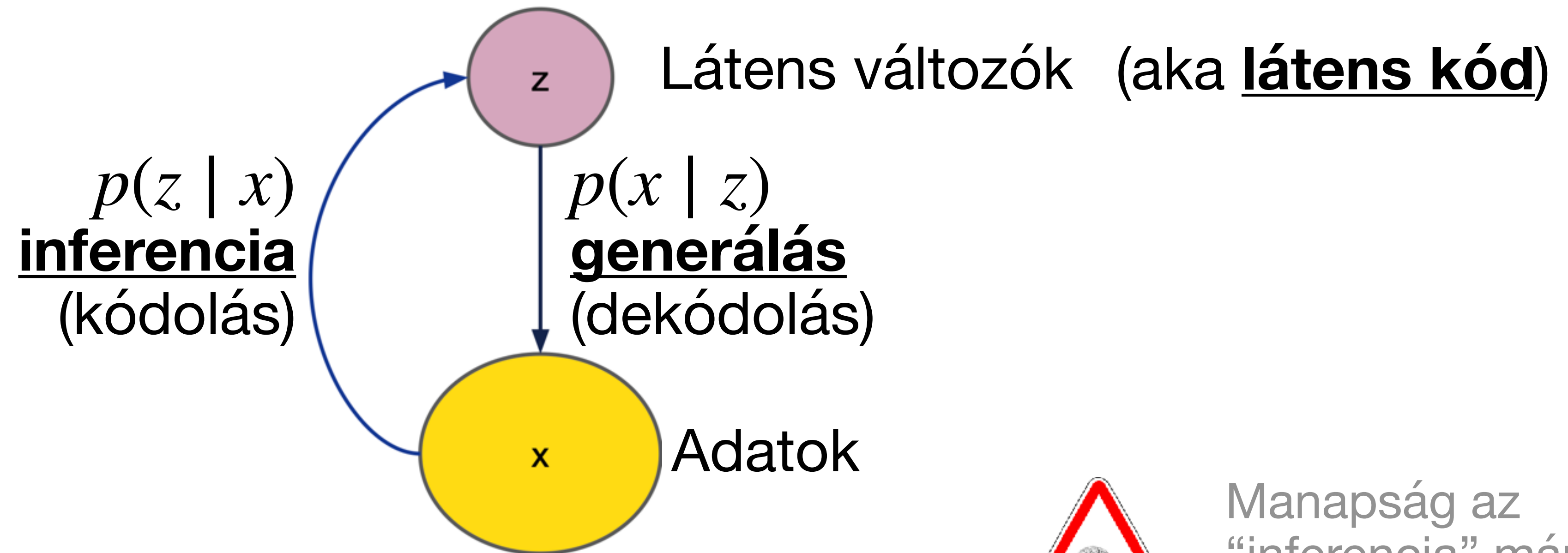


látens változók

Forrás: M. Rosca (Google DeepMind)

Generatív Modellelés

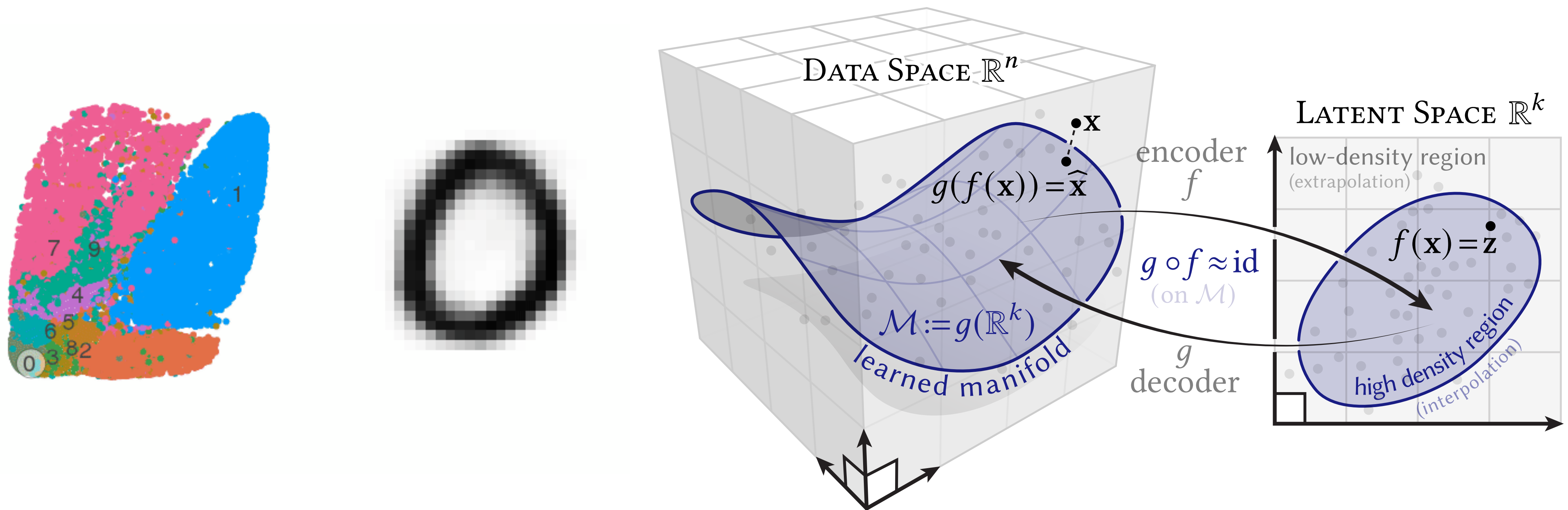
Látens változók



Manapság az “inferencia” már (gyakran) a *generálást* jelöli!

Generatív Modellelés

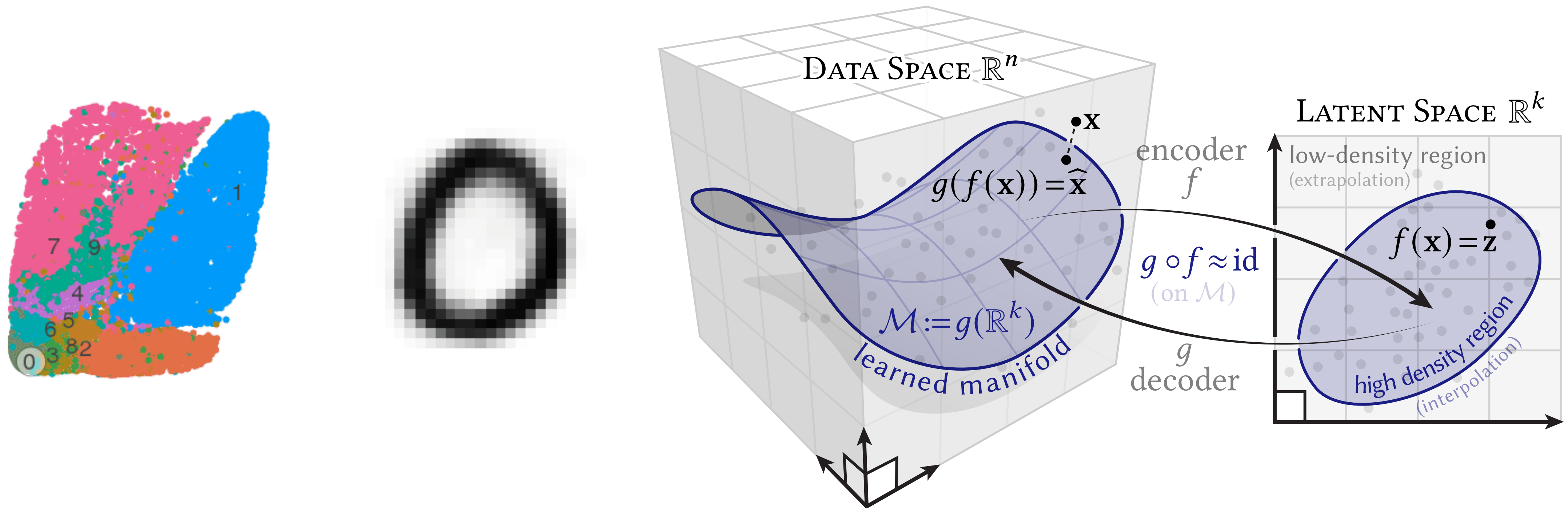
Látens tér



Geometriai értelmezés: dimenzió redukció — az (alacsonyabb dimenziójú) látens tér paraméterezi az adatsokaságot (data manifold)!

Generatív Modellelés

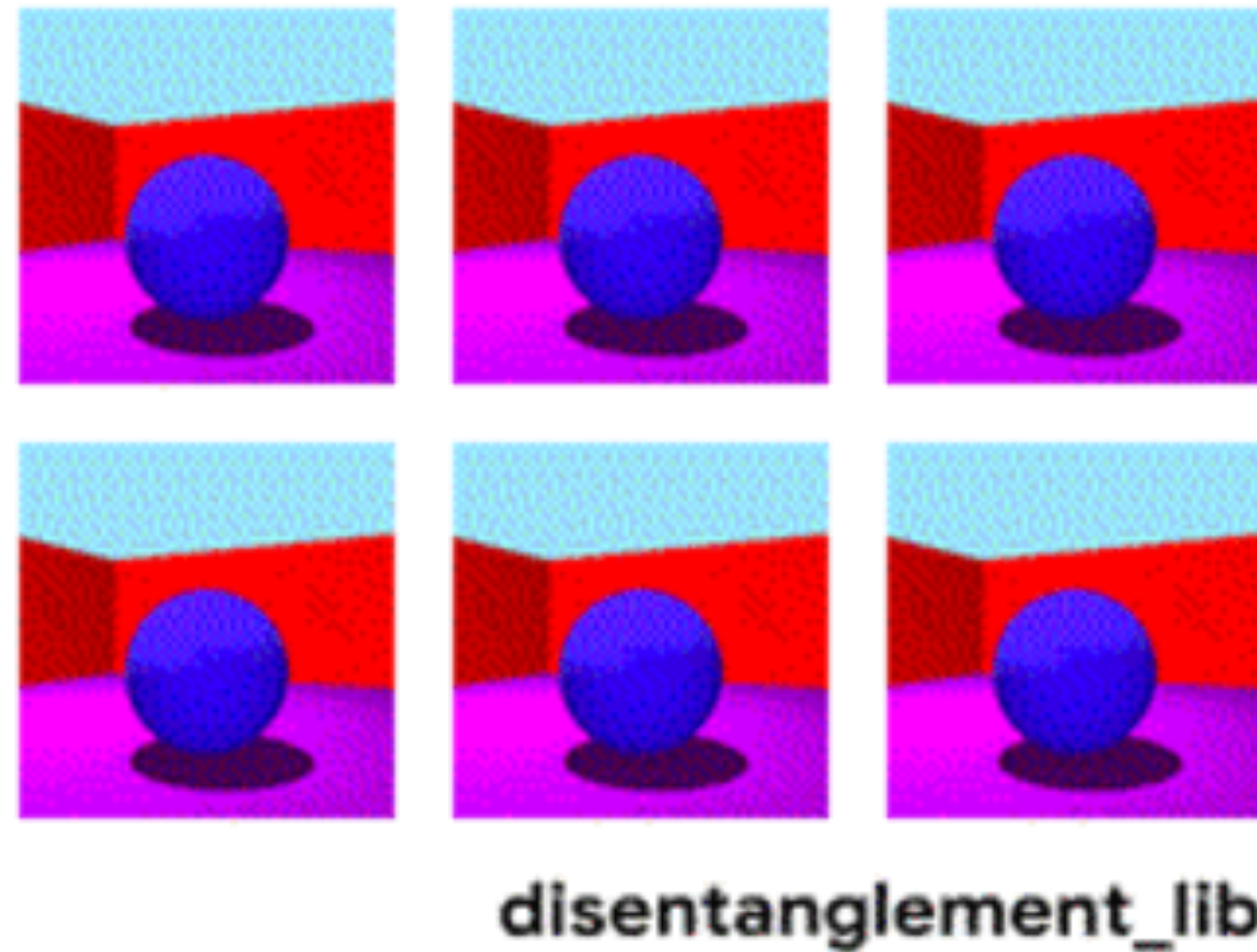
Látens tér



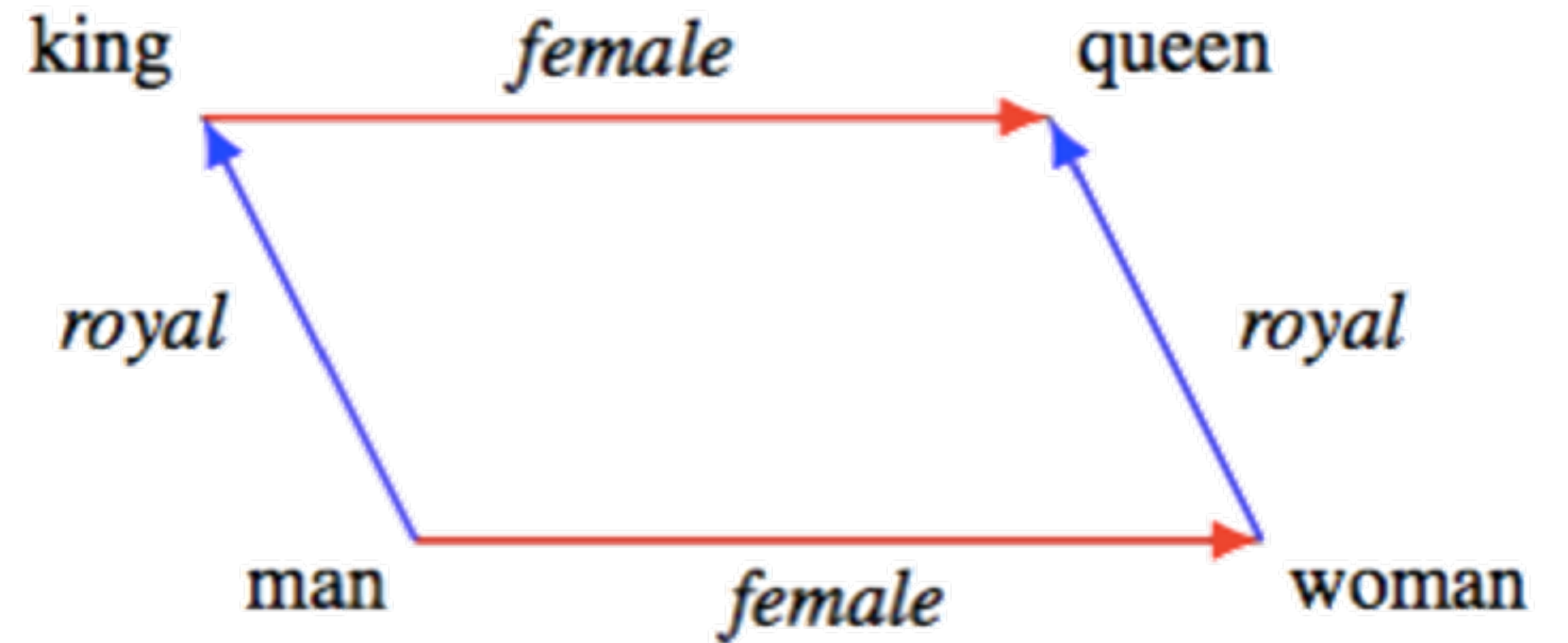
Geometriai értelmezés: dimenzió redukció — az (alacsonyabb dimenziójú) látens tér paraméterezi az adatsokaságot (data manifold)!

Generatív Modellelés

Látens tér – Kívánalmak



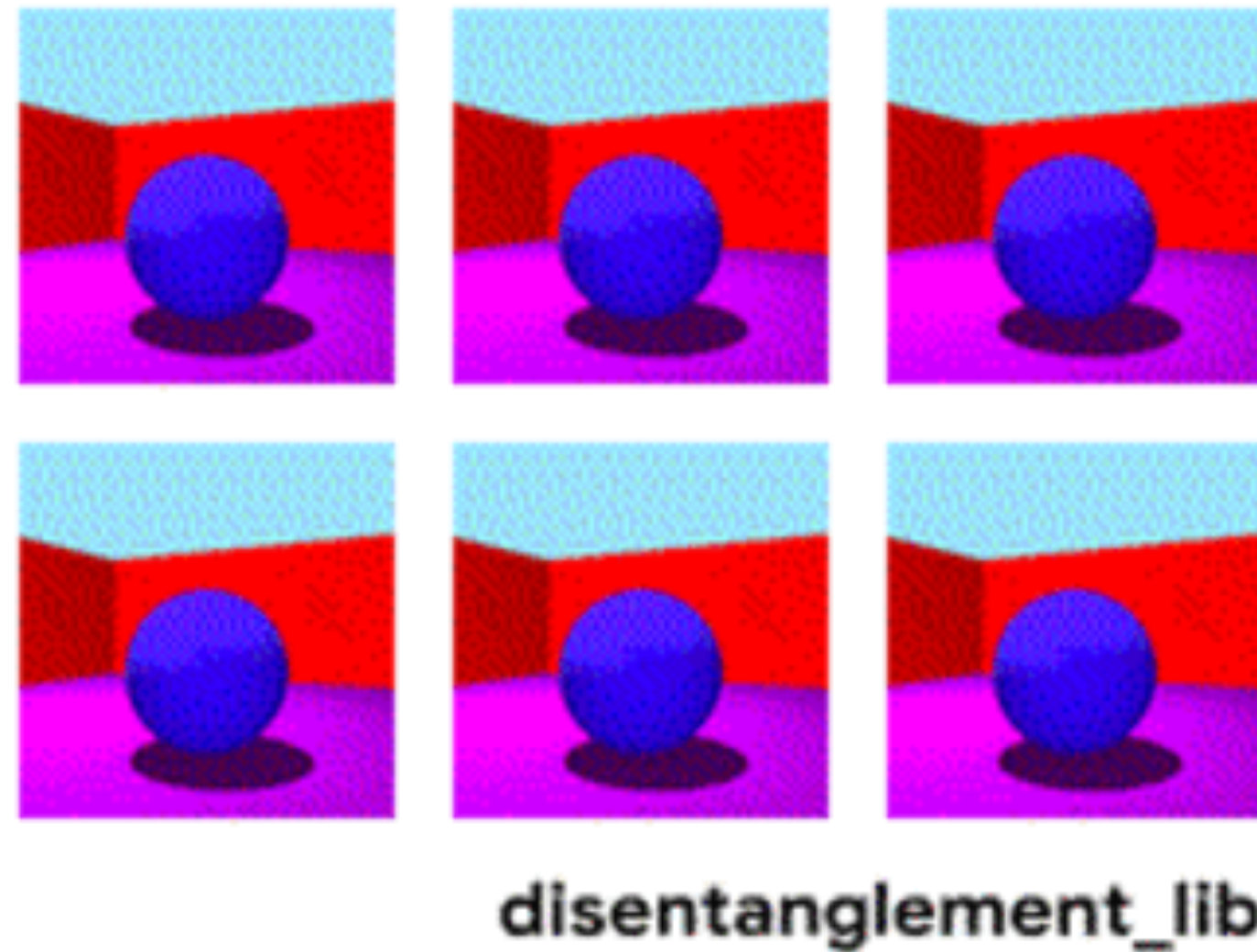
Szemantikai szétválaszthatóság



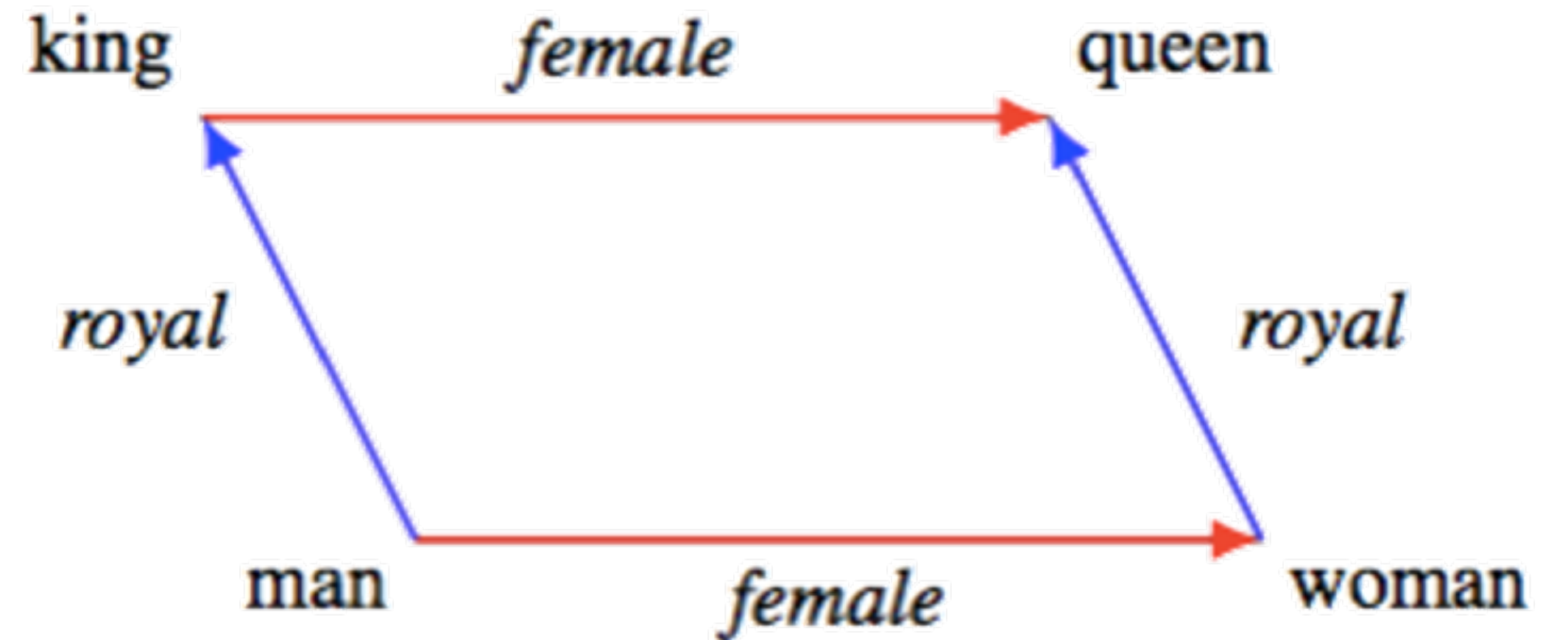
Linearitás (vektor aritmetika)

Generatív Modellelés

Látens tér – Kívánalmak



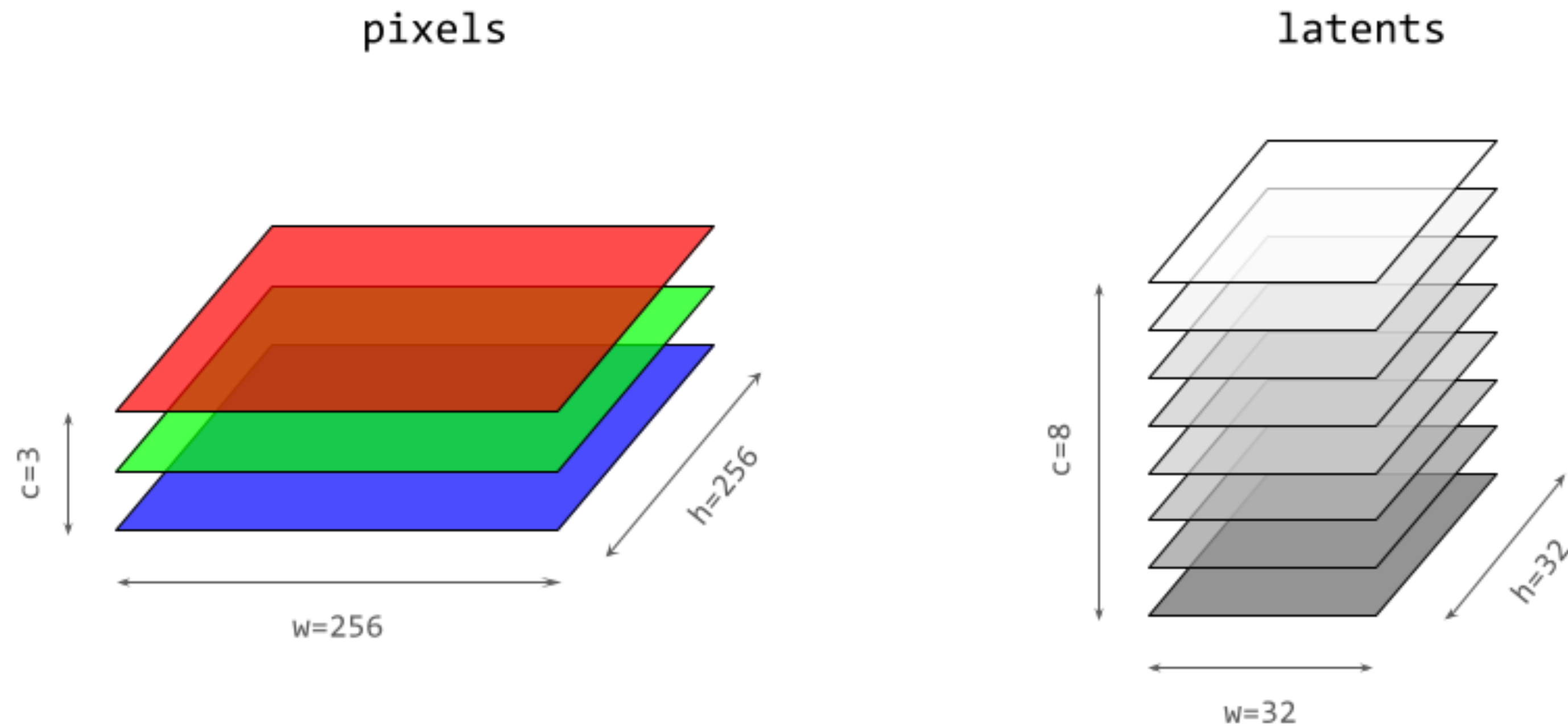
Szemantikai szétválaszthatóság



Linearitás (vektor aritmetika)

Generatív Modellezés

Látens tér – Képekre

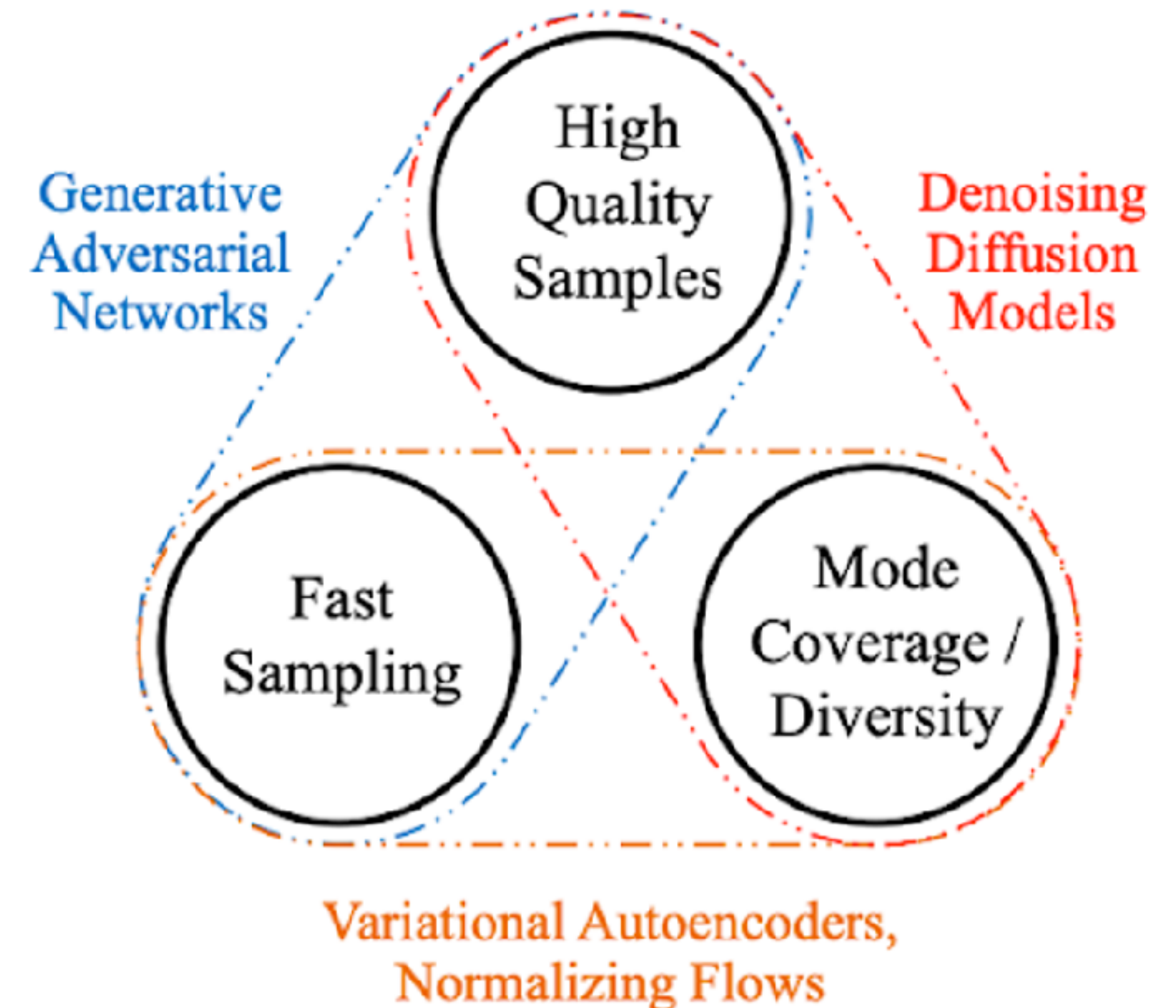


Képekre a látens gyakran maga is egy (kisebb méretű, sokcsatornás) kép (tenzor)!

Generatív Modellelés

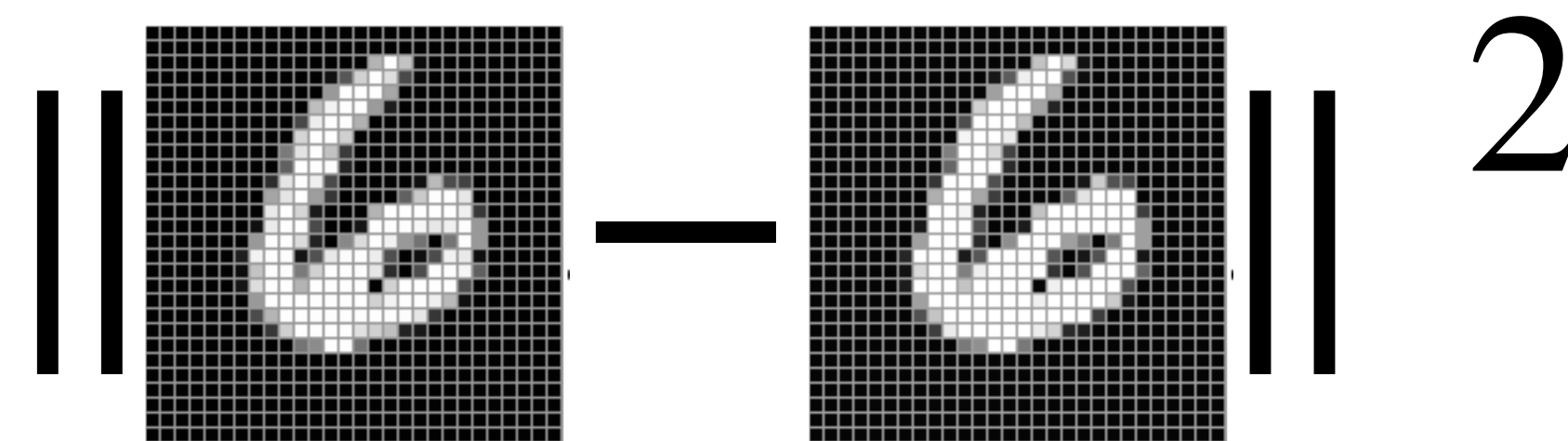
Elvárások

- A generatív modellekkel szemben támasztott elvárások:
 - A generált adatok legyenek “valódiak”, “érdekesek”, általában “jó minőségűek”!
 - A generált adatok legyenek sokszínűek, fedjék le a teljes adathalmazt, ne csak egy részhalmazát!
 - A generálás folyamata legyen minél gyorsabb, minél kevésbé erőforrásigényes!
- Egyszerre nehéz teljesíteni — “Trilemma”!



Képek Összehasonlítása

Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR)



$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Mean Error Squared

Átlagos négyzetes hiba



$$\begin{aligned} \text{PSNR} &= 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \\ &= 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \\ &= 20 \cdot \log_{10}(\text{MAX}_I) - 10 \cdot \log_{10}(\text{MSE}). \end{aligned}$$

Maximális jel-zaj arány [dB!]



Képek Összehasonlítása

Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR)



X



Y_1



Y_2

Melyik van közelebb MSE értelmében?

Képek Összehasonlítása

Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR)

$$MSE(Y_1, X) \leq MSE(Y_2, X)$$



X



Y_1



Y_2

Melyik van közelebb MSE értelmében?


Képek Összehasonlítása

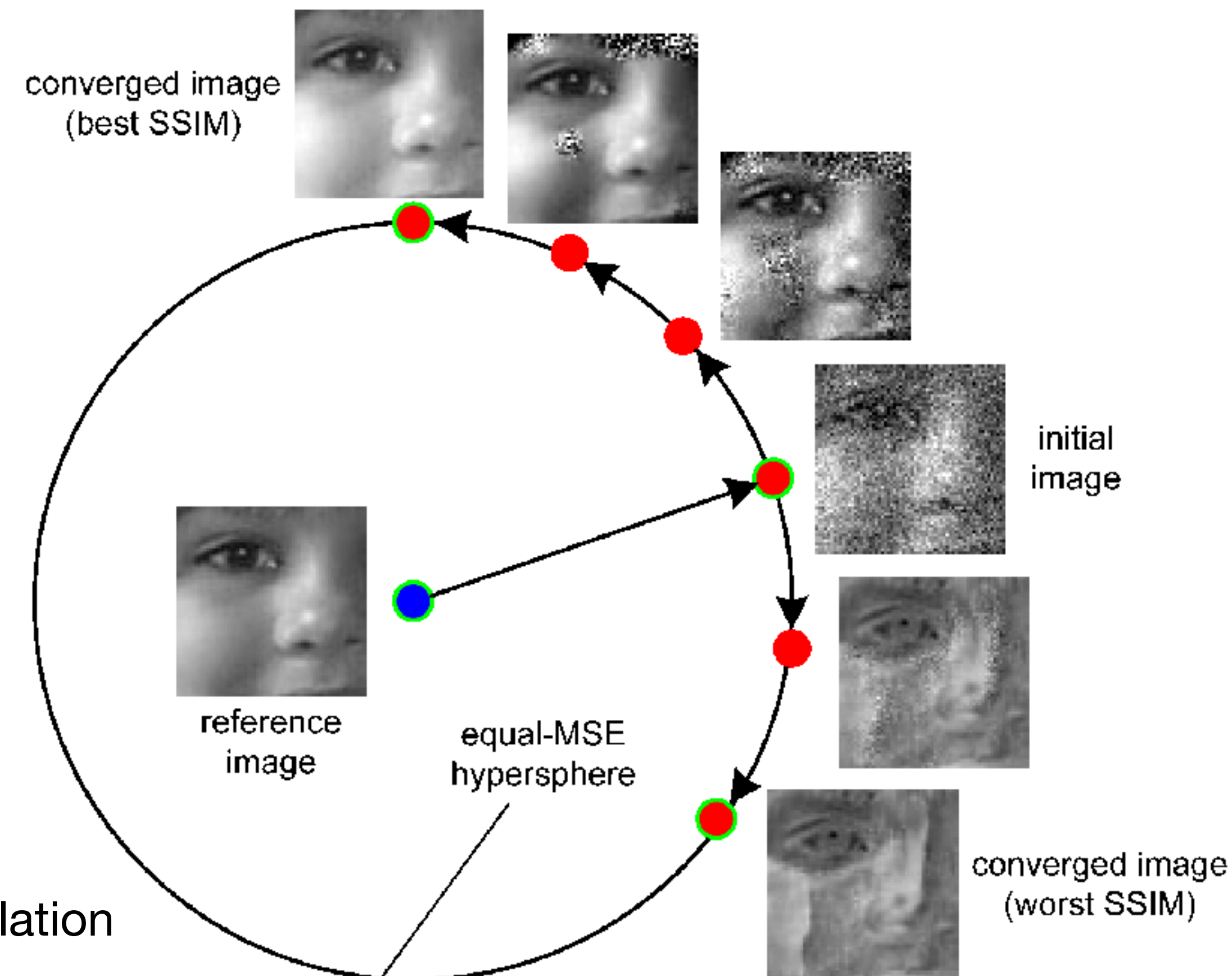
Structural Similarity Index Measure (SSIM)

$$SSIM(x, y) = l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma$$

luminancia kontraszt struktúra

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

$-1 \leq SSIM \leq 1$  $SSIM = \begin{cases} 1 : \text{perfect similarity} \\ 0 : \text{no similarity} \\ -1 : \text{perfect anti-correlation} \end{cases}$



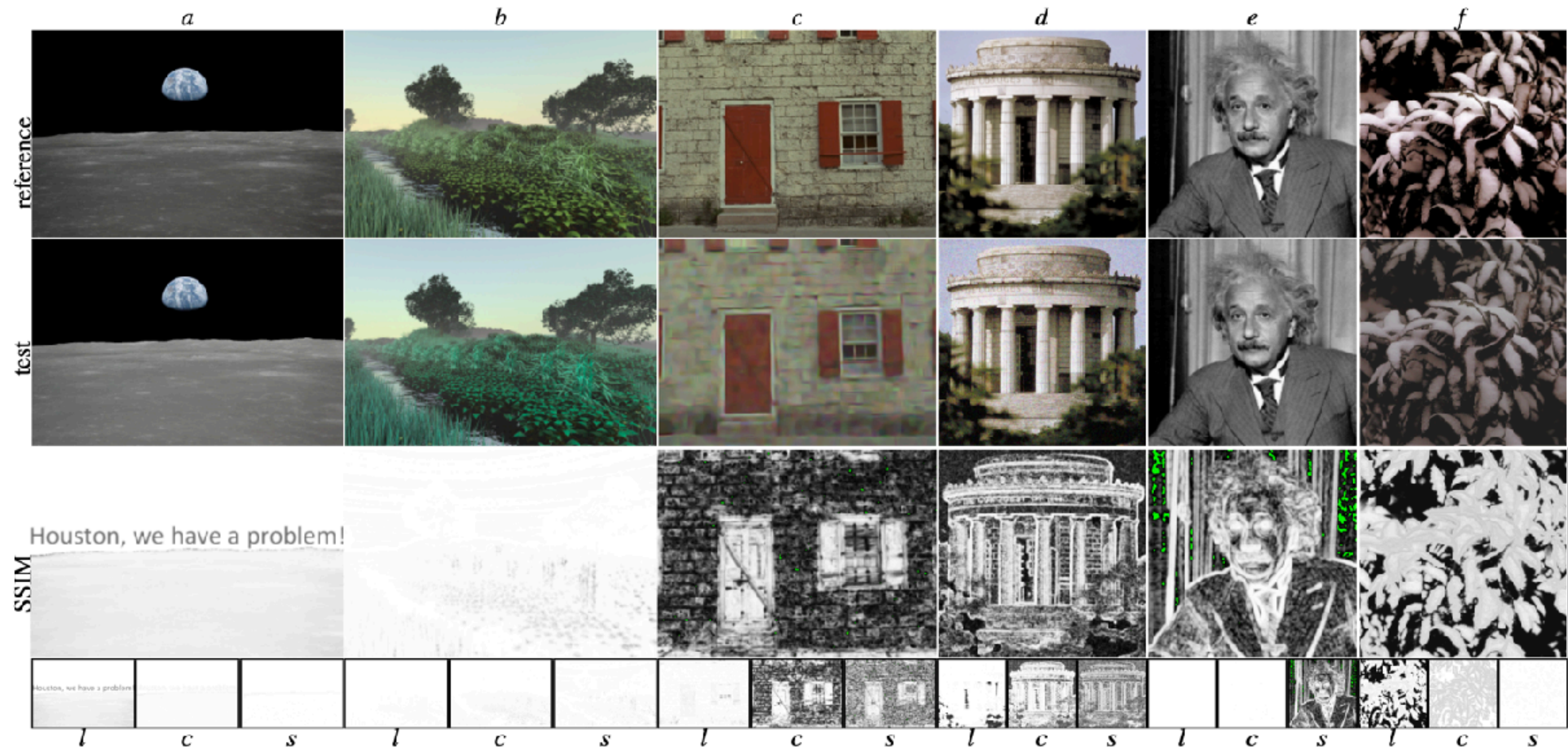
Képek Összehasonlítása

Structural Similarity Index Measure (SSIM)

Understanding SSIM

Jim Nilsson
NVIDIA

Tomas Akenine-Möller
NVIDIA



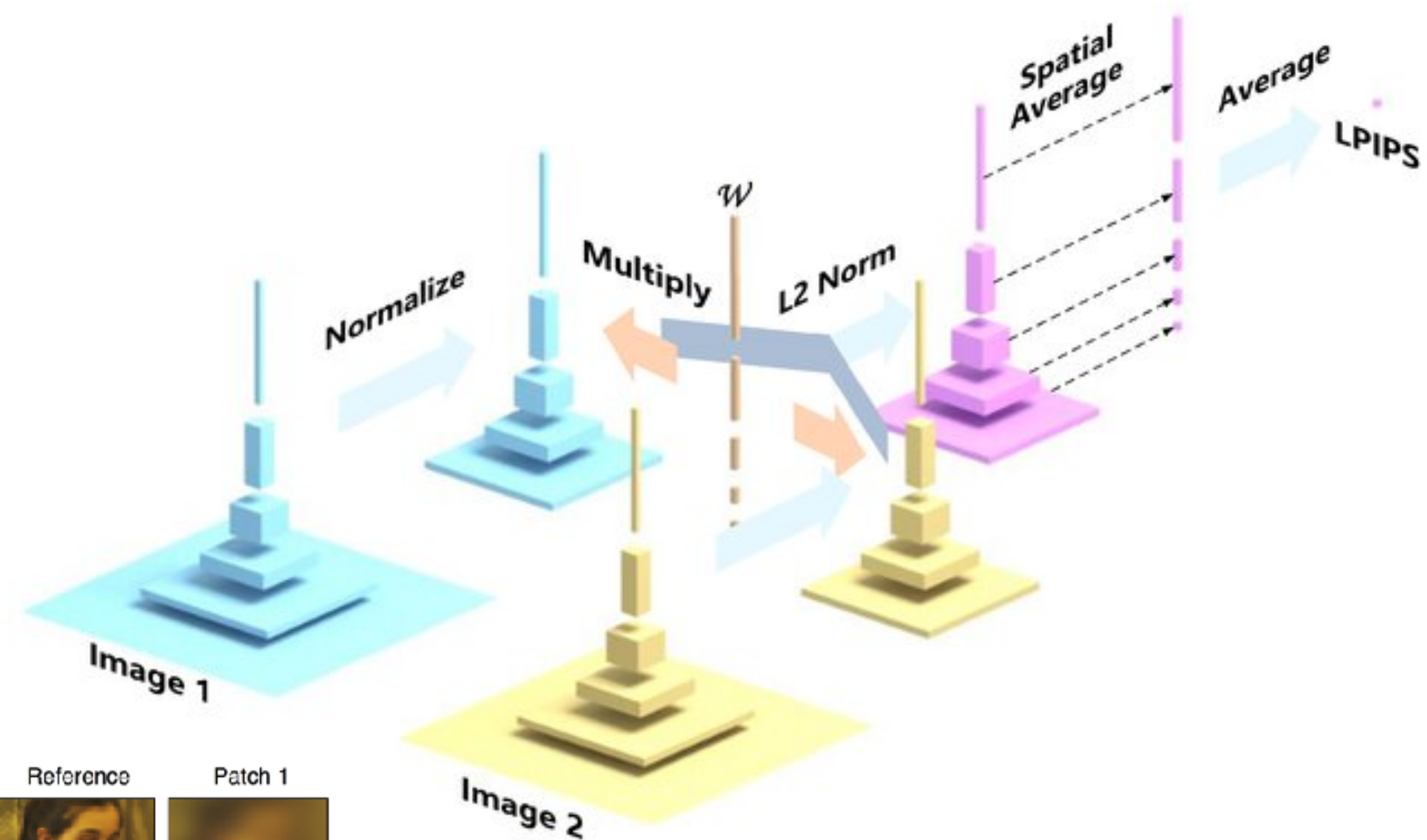
Képek Összehasonlítása

Learned Perceptual Image Patch Similarity (LPIPS)

The Unreasonable Effectiveness of Deep Features as a Perceptual Metric

Richard Zhang¹ Phillip Isola^{1,2} Alexei A. Efros¹
¹UC Berkeley ²OpenAI

Eli Shechtman³ Oliver Wang³
³Adobe Research



Humans
 L2/PSNR, SSIM, FSIM
 Random Networks
 Unsupervised Networks
 Self-Supervised Networks
 Supervised Networks

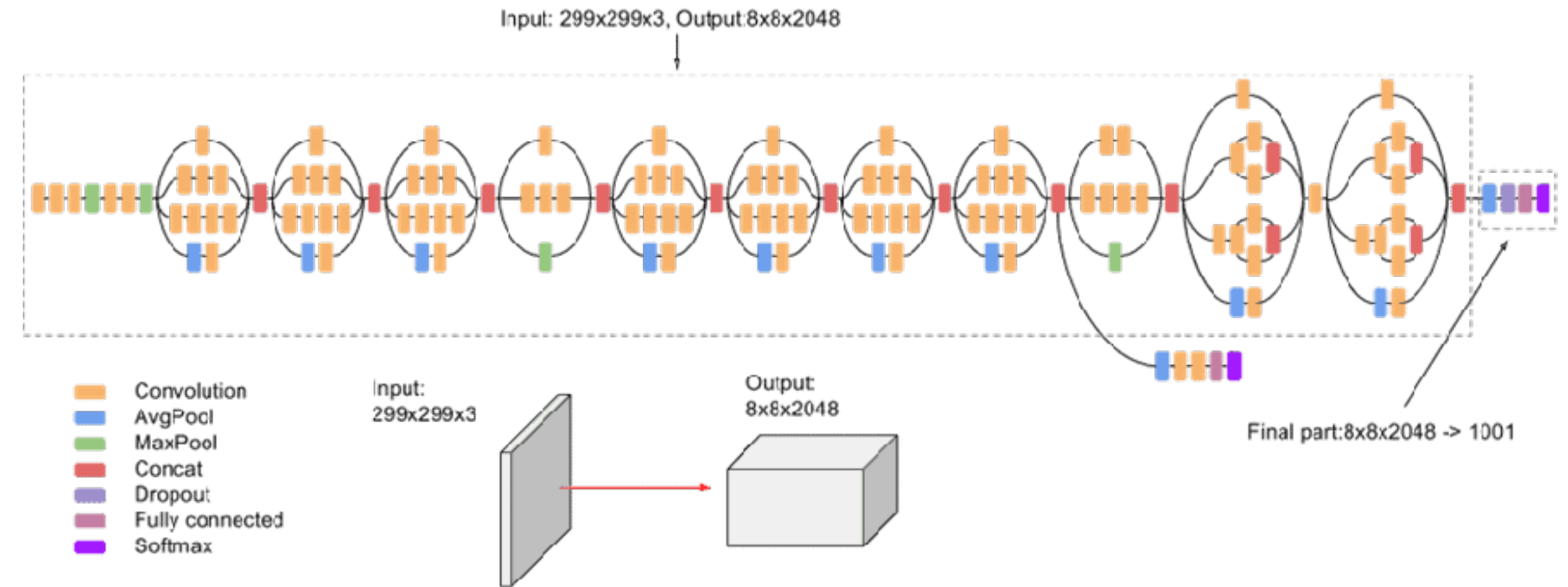
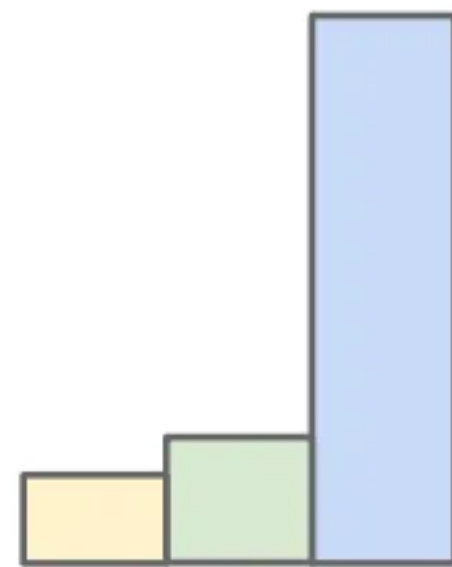


Generatív Modellek Minősítése

Inception Score (IS)



Inception v3



Inception v3

1 képre: ImageNet címkék eloszlása legyen “csúcsos”



Inception v3



N képre: ImageNet címkék eloszlása legyen “egyenletes”

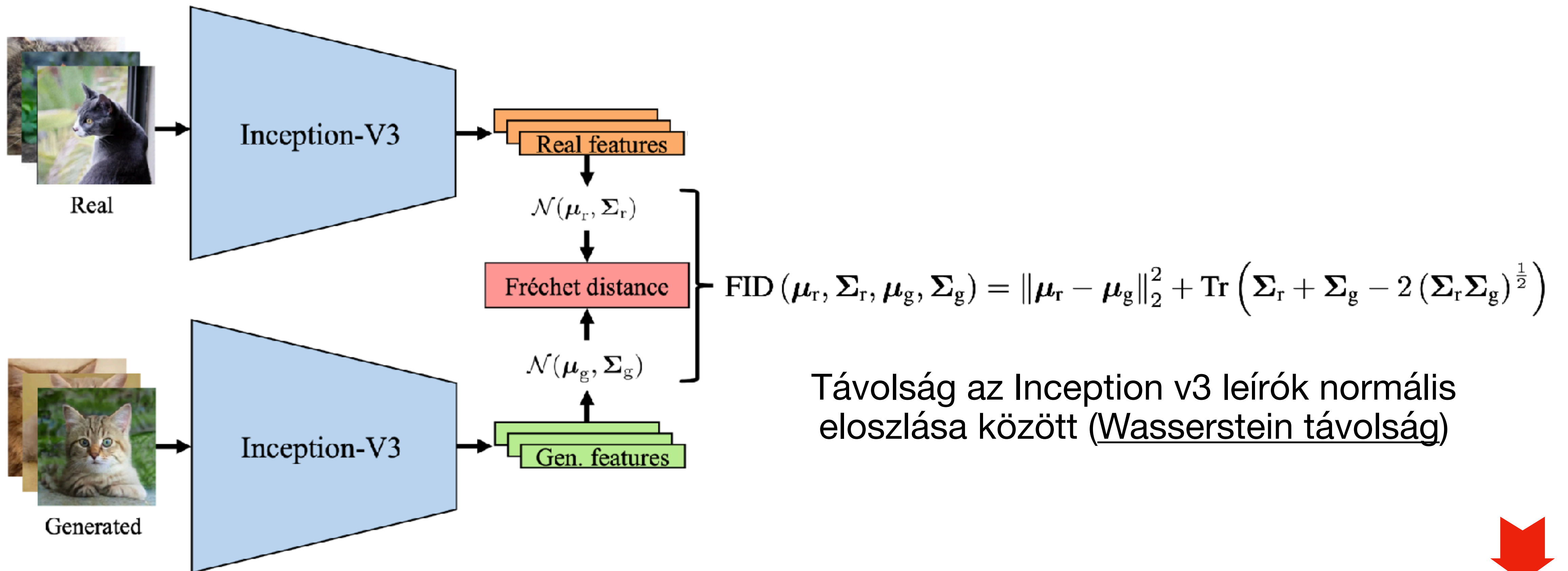
KL divergencia a címkék marginális & képenkénti feltételes eloszlása között:

$$\begin{aligned}
 IS(x) &= \exp(\mathbb{E}_{x \sim p_g} [D_{KL} [p(y|x) \parallel p(y)]]) \\
 &= \exp(\mathbb{E}_{x \sim p_g, y \sim p(y|x)} [\log p(y|x) - \log p(y)]) \\
 &= \exp(H(y) - H(y|x))
 \end{aligned}$$



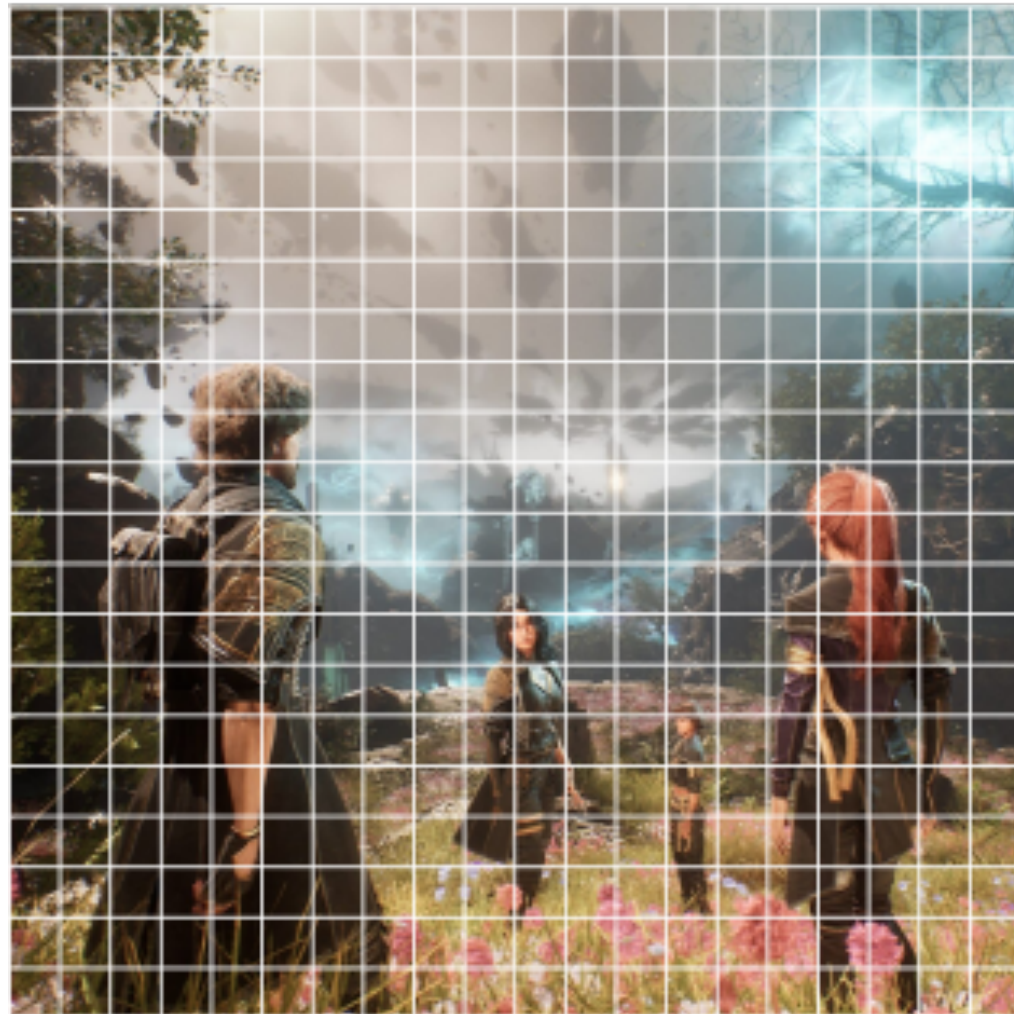
Generatív Modellek Minősítése

Frechet Inception Distance (FID)



Autoenkóder

Input



Encoder



Decoder



Z
0.1
1.4
5.1
0.3

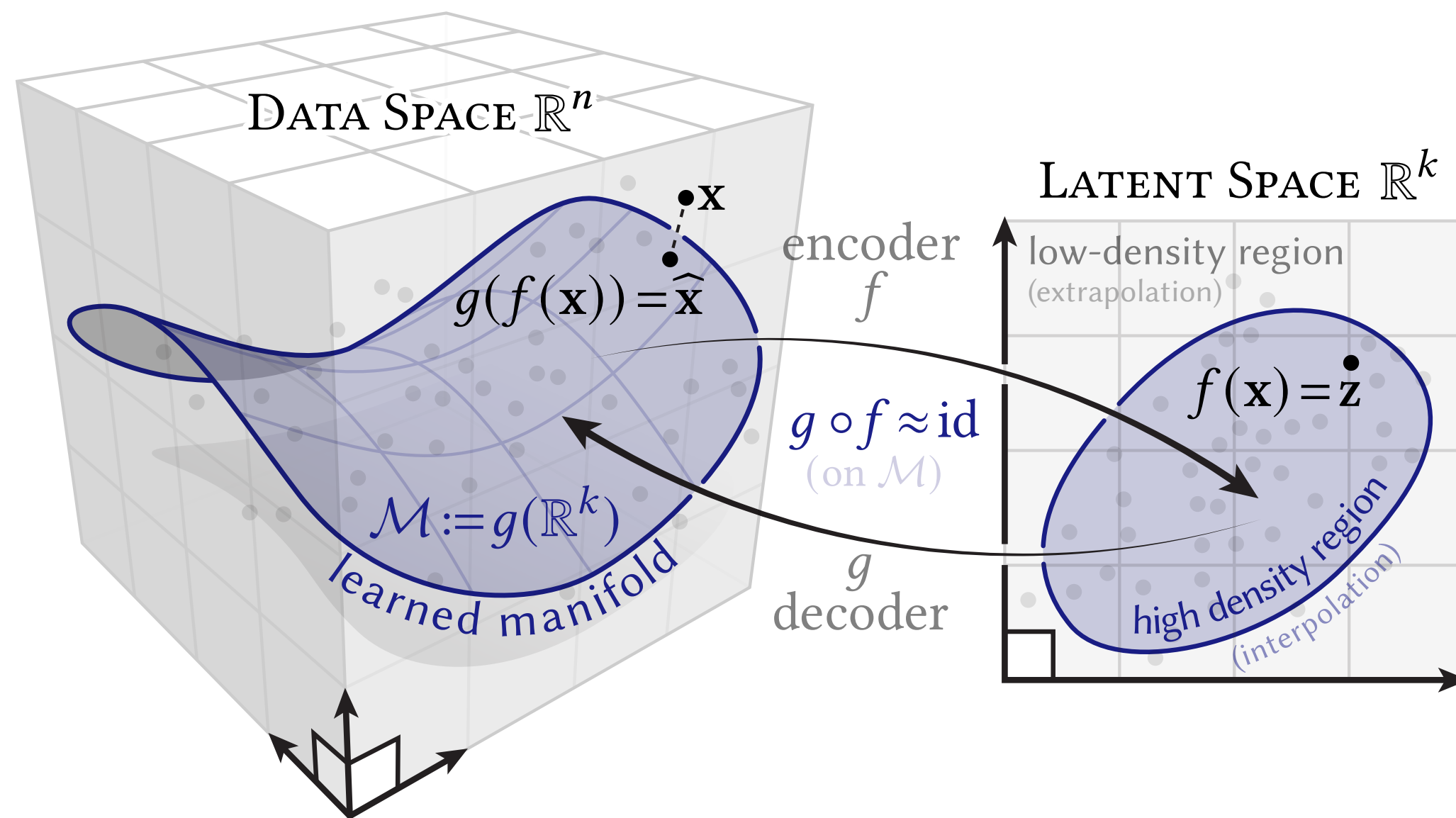
Reconstruction



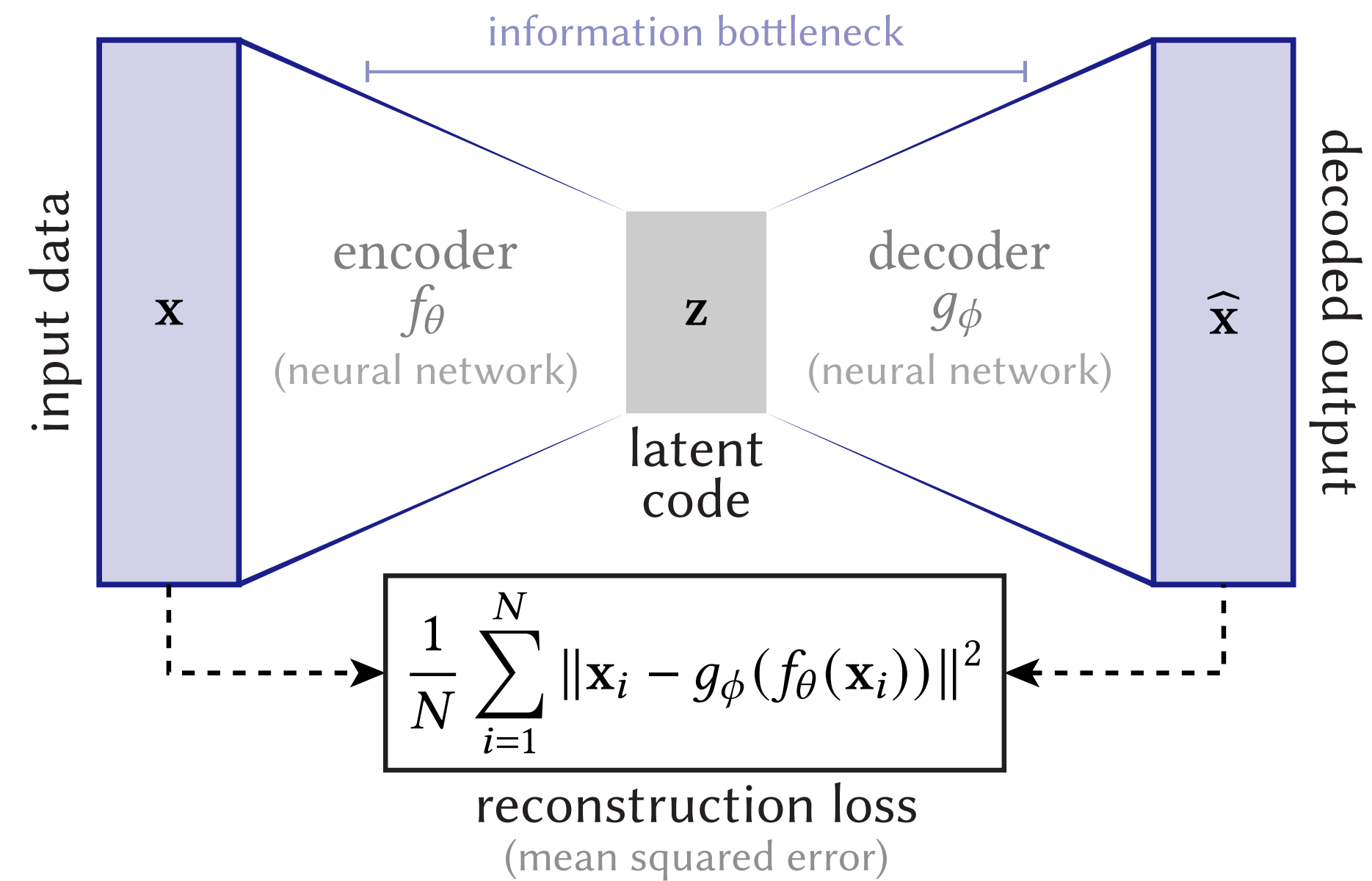
Autoenkóder

Autoencoder

What does it represent?



How is it implemented?



Autoenkóder

Lineáris autoenkóder – PCA

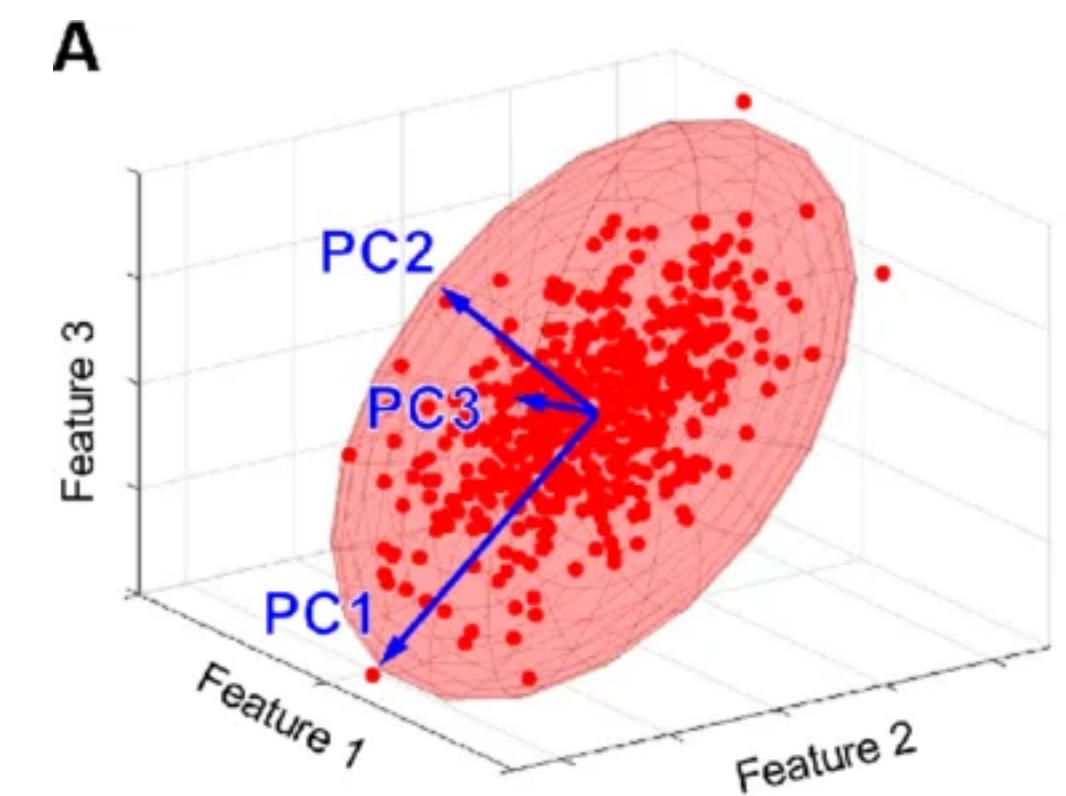
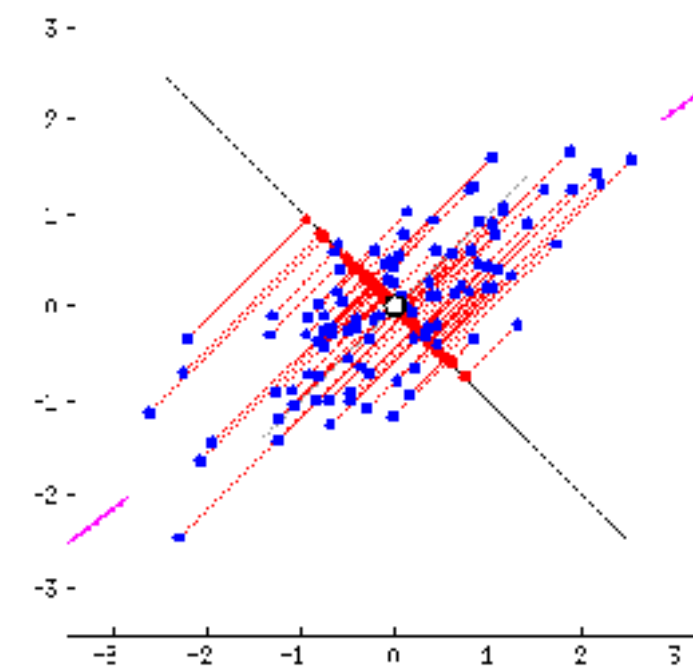
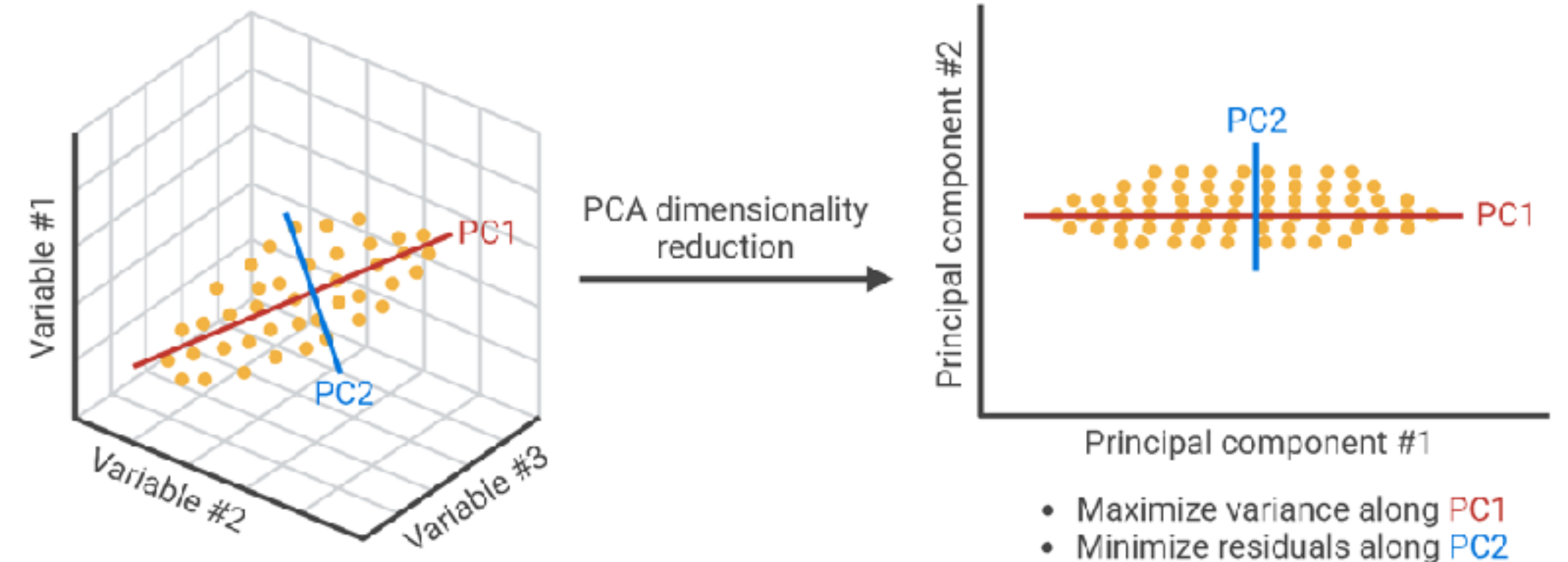
- Mely irányok mentén legnagyobb/legkisebb az adatok varianciája?
- Képezzük az adatpontok (empirikus) kovariancia-mátrixát:

$$\mathbf{Q} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = (\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})^T$$

- Számoljuk ki \mathbf{Q} sajátértékeit & sajátvektorokat (azaz \mathbf{X} szinguláris érték felbontását (SVD))

$$\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}^2\mathbf{V}^T = \sum_{i=1}^N \sigma_i^2 \mathbf{u}\mathbf{v}^T$$

- Kiválasztjuk a k db legnagyobb σ_i -hez tartozó vektort (főkomponenst) – **Principal Component Analysis (PCA)**
- A PCA (aka “faktor analízis”) további értelmezései:
 - az adatok legjobb k -dim. *lineáris* modellje (távolság² szerint)
 - az adatokra legjobb illeszkedő k -dim. *ellipszoid (normális eloszlás)*
 - a mátrix közelítése a legjobb k -rangú mátrix-szal



Autoenkóder

Lineáris autoenkóder – PCA

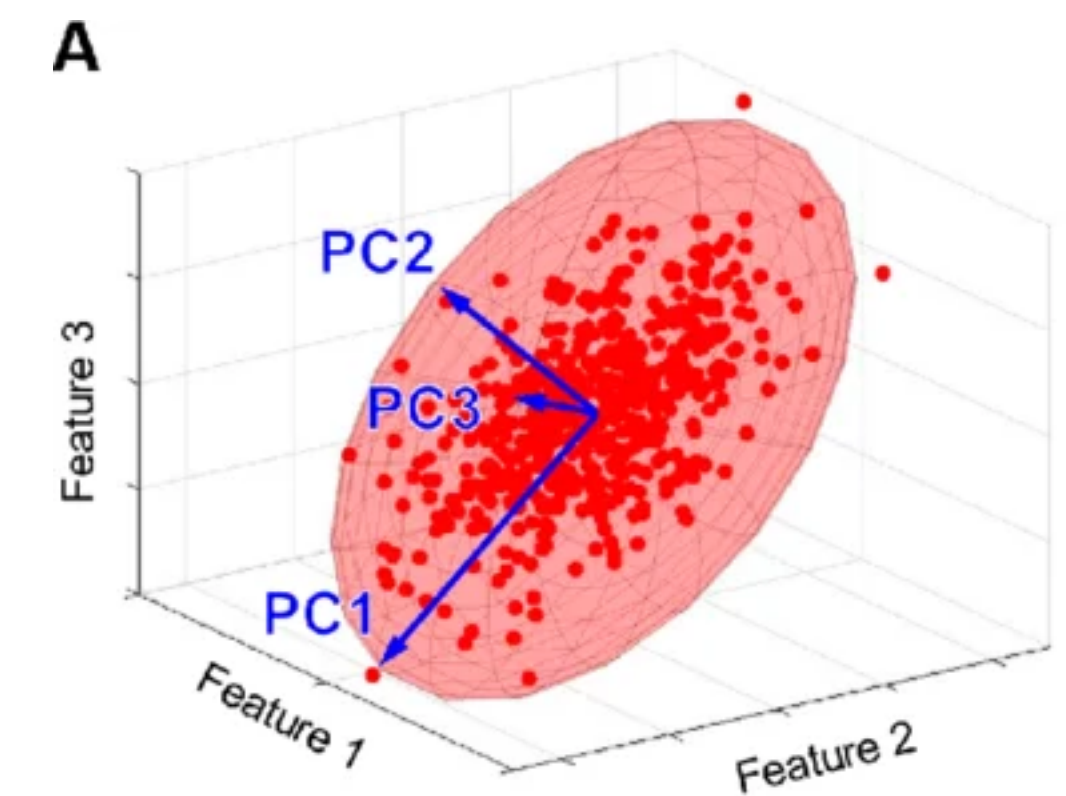
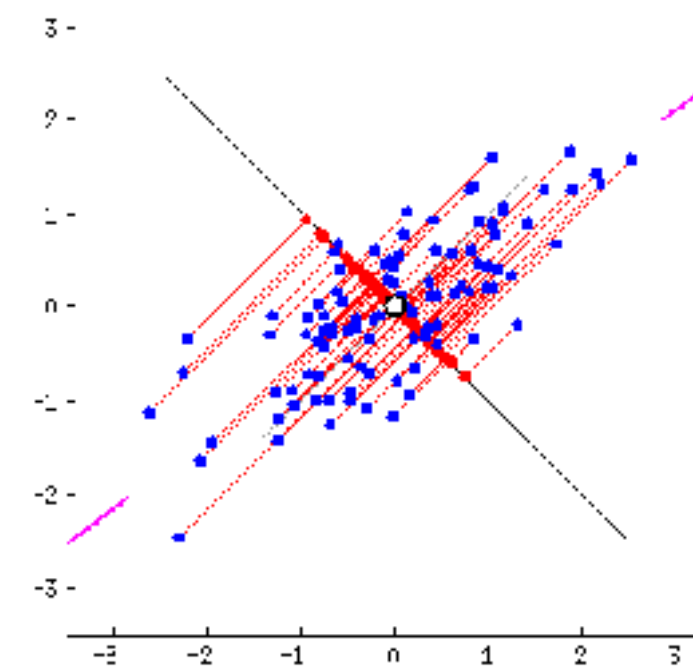
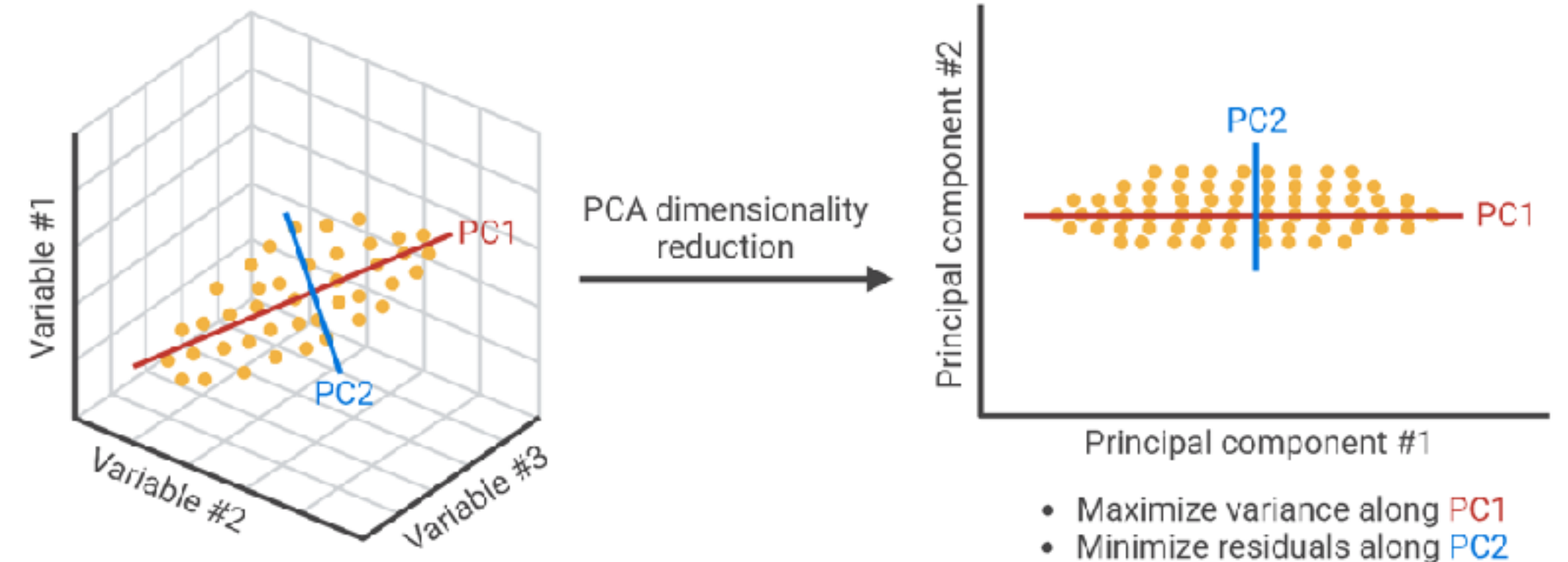
- Mely irányok mentén legnagyobb/legkisebb az adatok varianciája?
- Képezzük az adatpontok (empirikus) kovariancia-mátrixát:

$$\mathbf{Q} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = (\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})(\mathbf{X} - \bar{\mathbf{x}}\mathbf{1})^T$$

- Számoljuk ki \mathbf{Q} sajátértékeit & sajátvektorokat (azaz \mathbf{X} szinguláris érték felbontását (SVD))

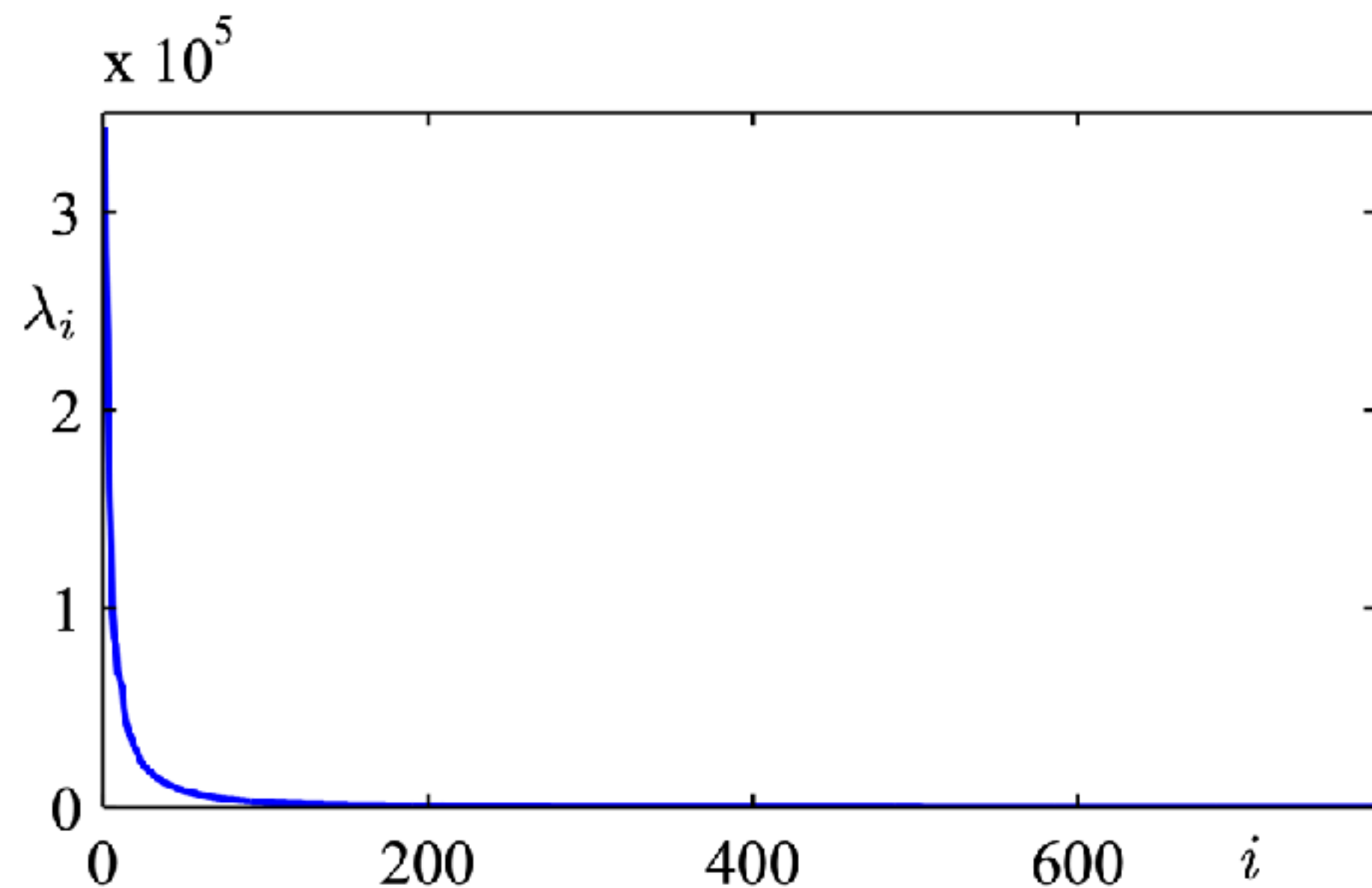
$$\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}^2\mathbf{V}^T = \sum_{i=1}^N \sigma_i^2 \mathbf{u}\mathbf{v}^T$$

- Kiválasztjuk a k db legnagyobb σ_i -hez tartozó vektort (főkomponenst) – **Principal Component Analysis (PCA)**
- A PCA (aka “faktor analízis”) további értelmezései:
 - az adatok legjobb k -dim. *lineáris* modellje (távolság² szerint)
 - az adatokra legjobb illeszkedő k -dim. *ellipszoid* (*normális eloszlás*)
 - a mátrix közelítése a legjobb k -rangú mátrix-szal

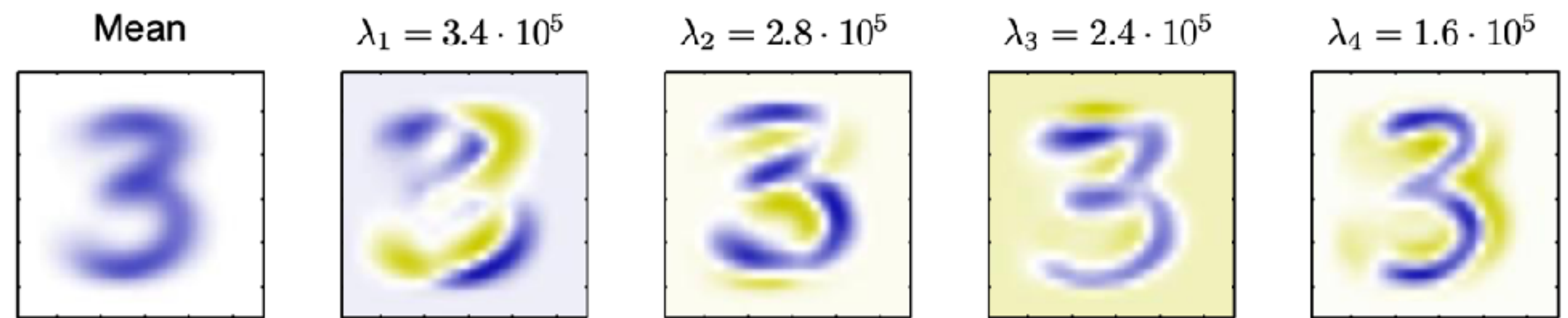


Autoenkóder

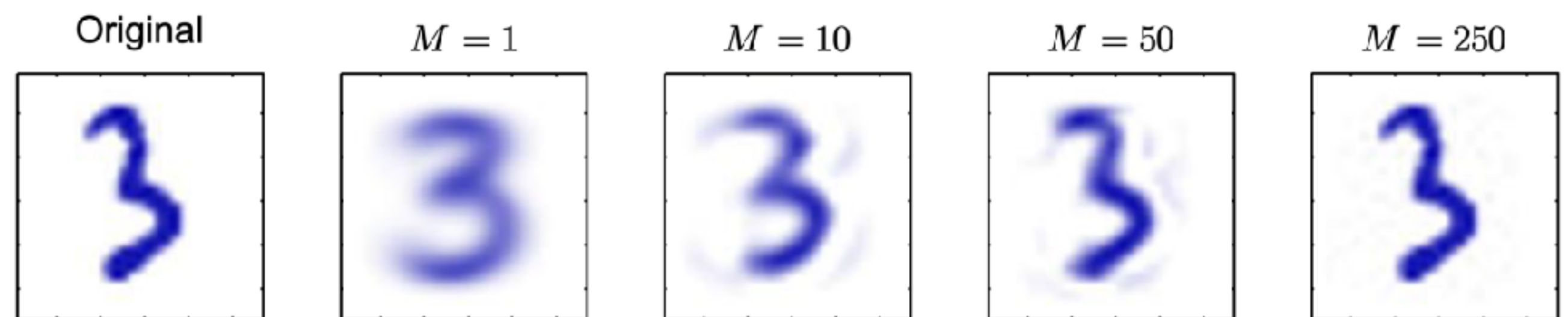
Lineáris autoenkóder – PCA – Példa (MNIST '3')



Kovarianciamátrix sajátértékei



Főkomponens vektorok

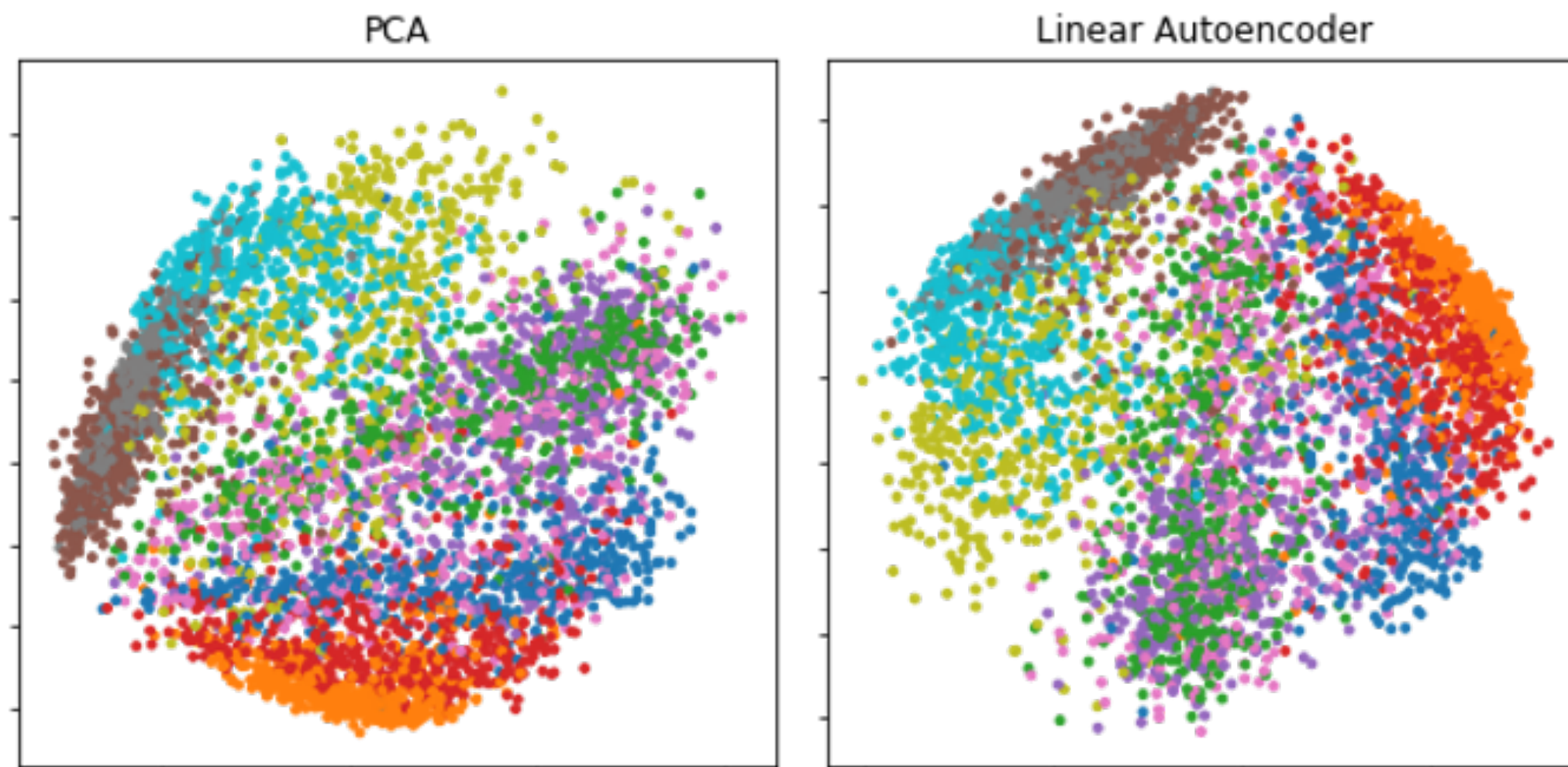


Rekonstrukció M főkomponenssel

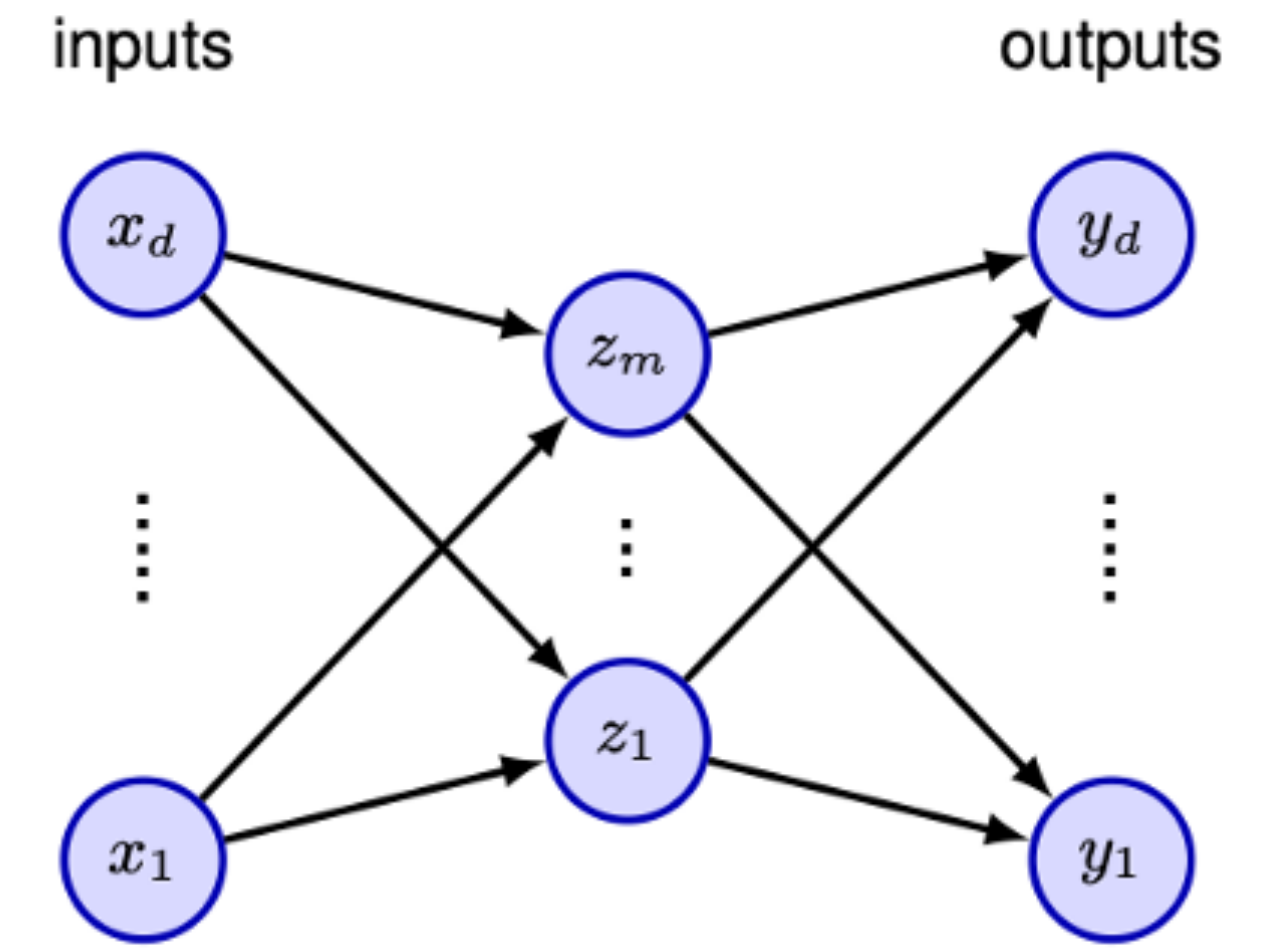
Autoenkóder

Mély autoenkóder

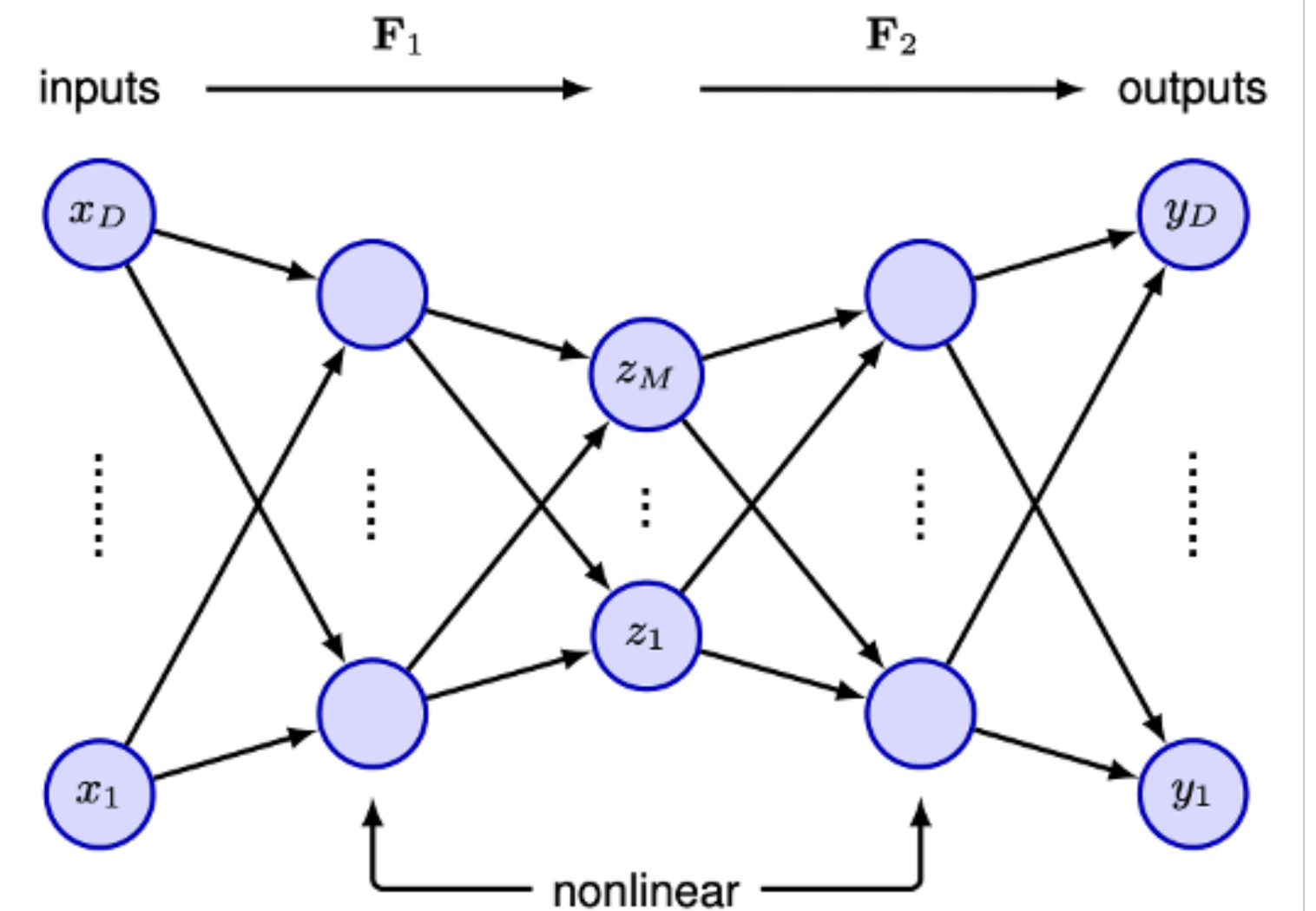
- Lineáris enkóder + dekóder rétegek MSE rekonstrukciós hibára tanítva \cong PCA! (modulo bázis transzformáció)



- Általánosítás: több réteg + nemlinearitások a rétegek között — mély autoenkóder neurális háló!



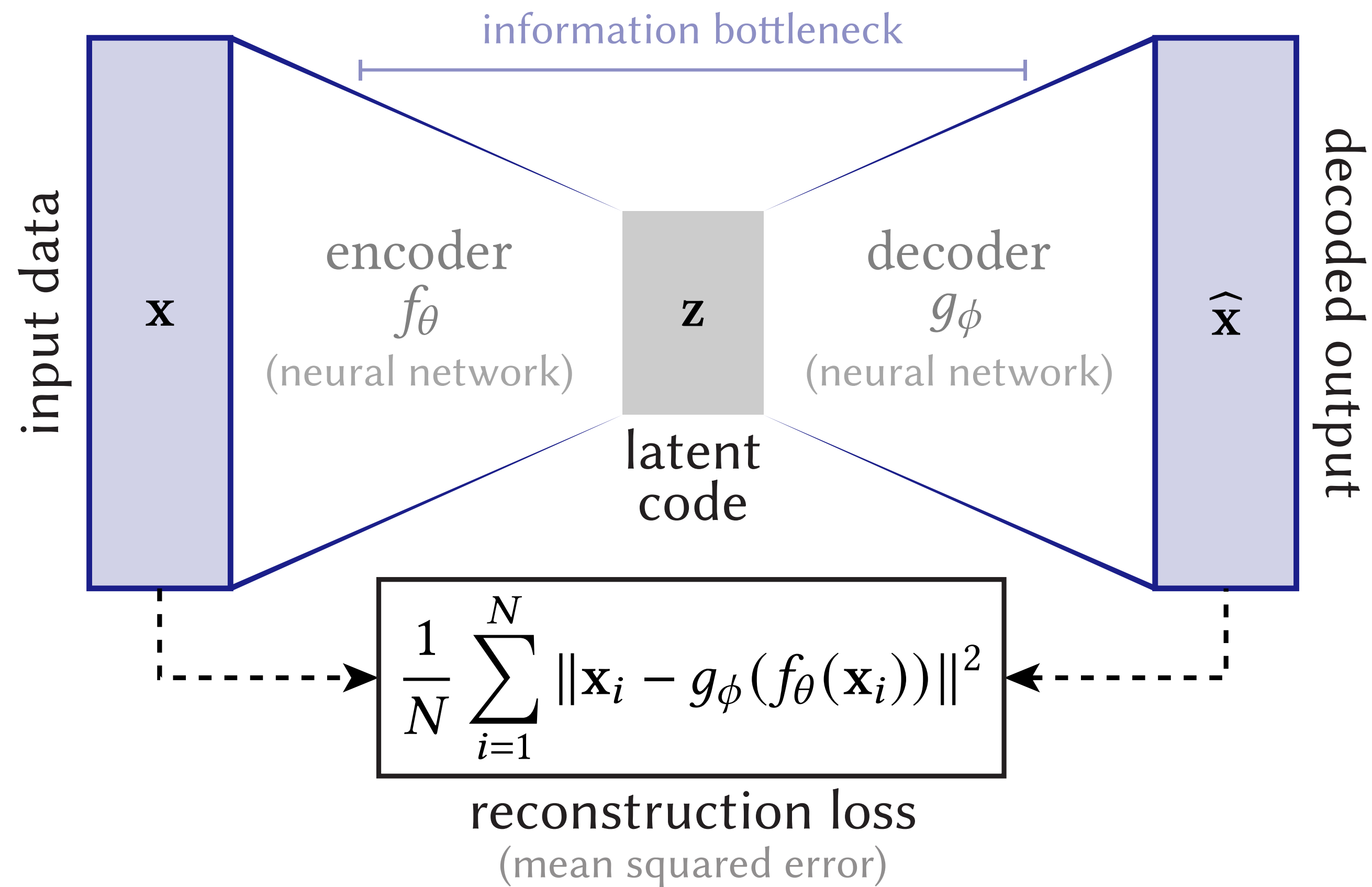
Lineáris Autoenkóder / PCA



Mély Autoenkóder

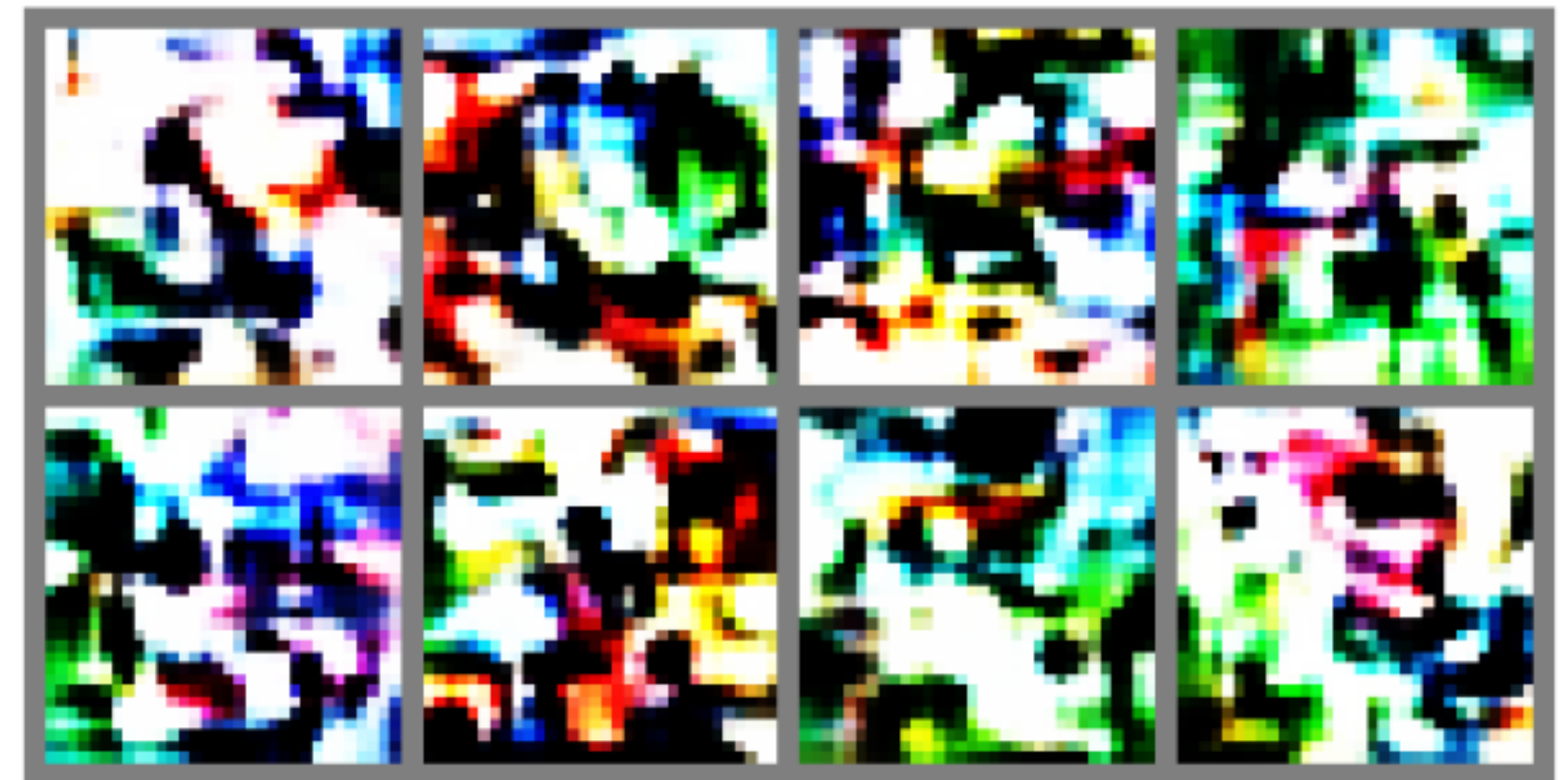
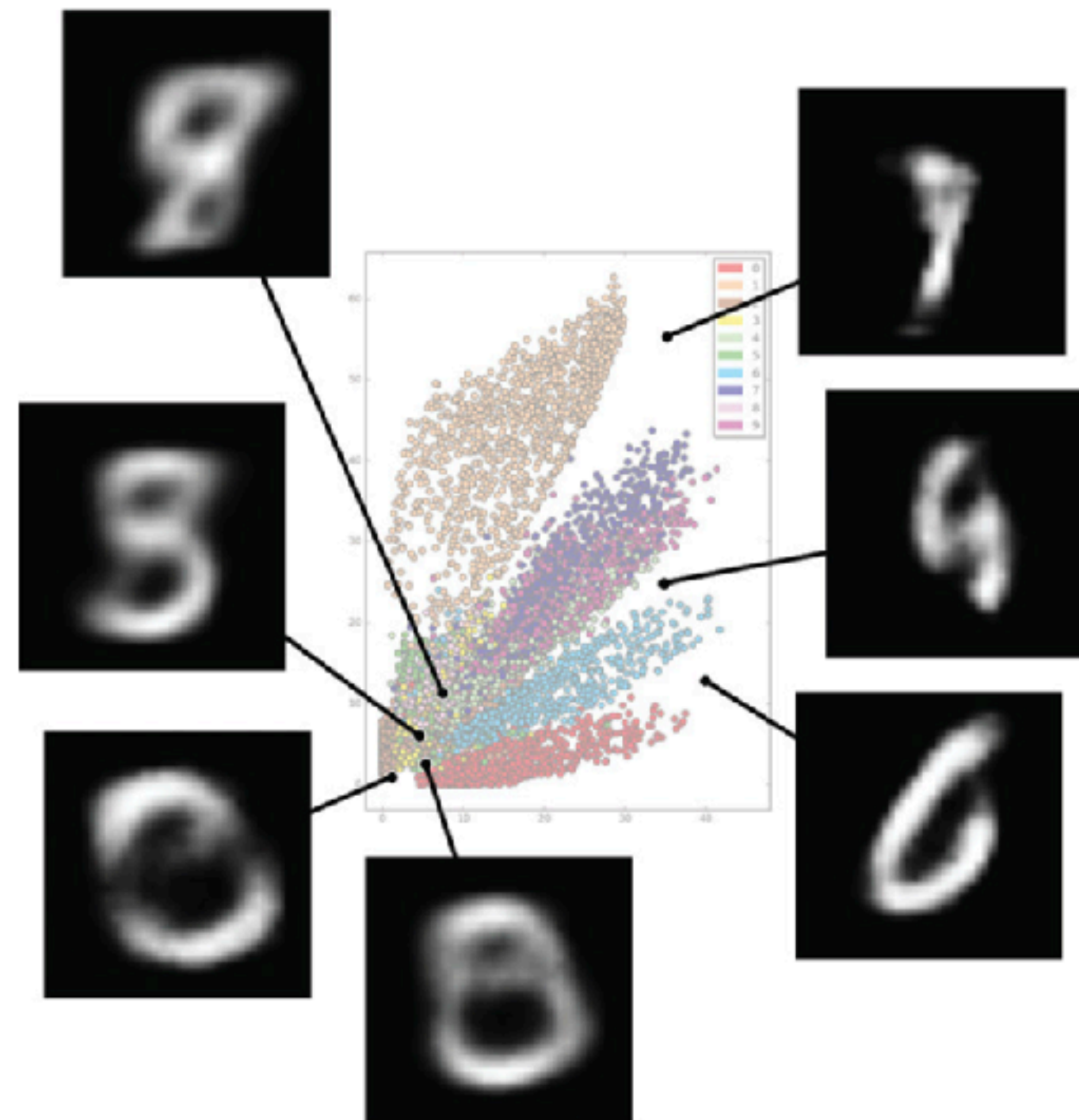
Autoenkóder

Mély autoenkóder



Autoenkóder

Mély autoenkóder – miért nem elég?



random látens kód CIFAR-10 autoenkóderrel dekódolva

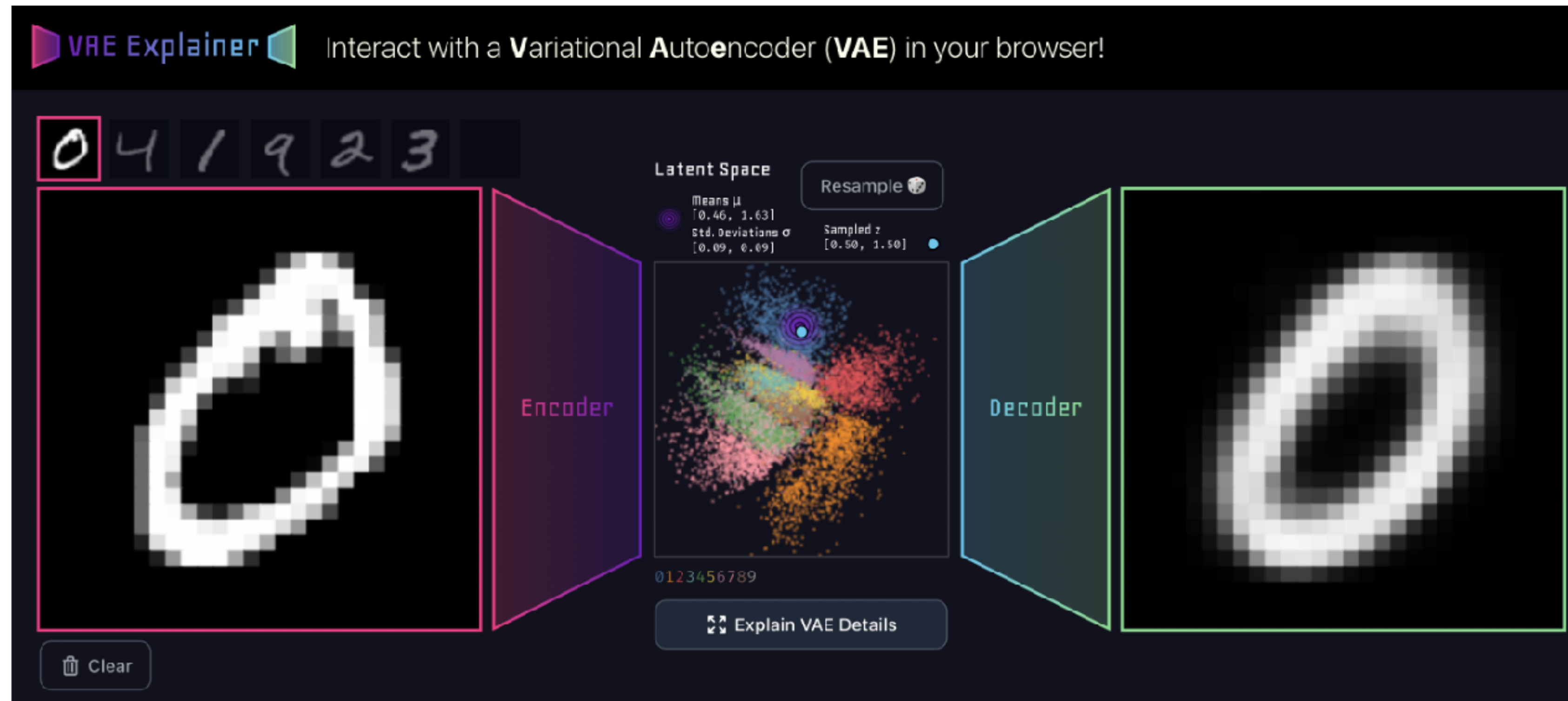
Ha csak rekonstrukcióra optimalizáljuk, a mély autoenkóder látens teréből nem lehet bármit dekódolni!
Valahogy regularizálni kell a látens tér eloszlását...

Variációs Autoenkóder (VAE)

Auto-Encoding Variational Bayes

Diederik P. Kingma
Machine Learning Group
Universiteit van Amsterdam

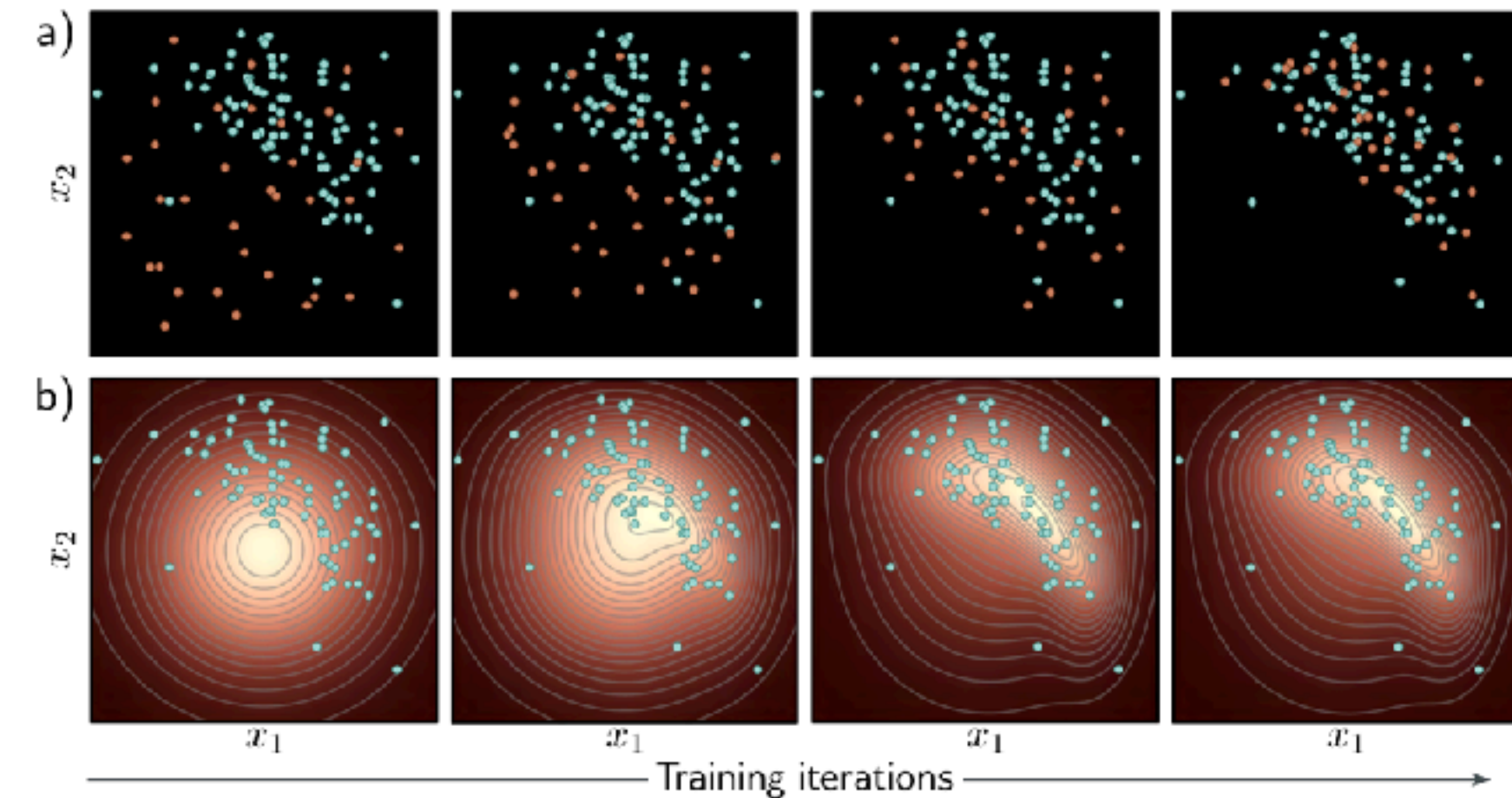
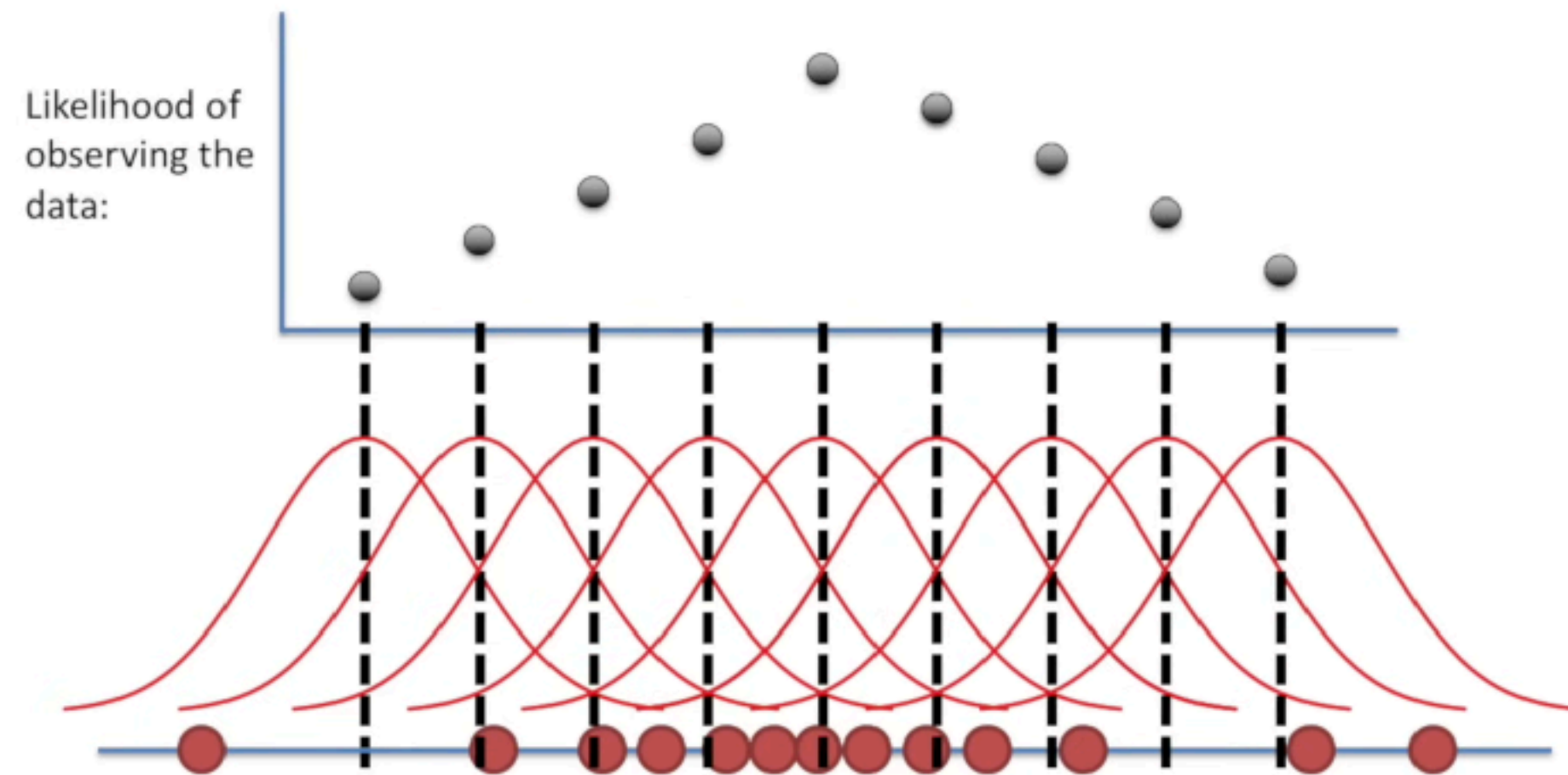
Max Welling
Machine Learning Group
Universiteit van Amsterdam



<https://xnought.github.io/vae-explainer/>

Variációs Autoenkóder (VAE)

Maximum Likelihood



$$\min_{p(x)} \text{KL}(p_{model}(x) || p_{true}(x)) \Leftrightarrow \min \sum_i p_{true}(x_i) \cdot \log p_{model}(x) - \overbrace{\sum p_{true}(x_i) \cdot \log p_{true}(x)}^{const!}$$
$$\Leftrightarrow \max \mathbb{E}_{p_{true}(x)}[\log p_{model}(x)]$$

Cél: maximalizálni az adatpontok valószínűségét
“maximum likelihood”

Variációs Autoenkóder (VAE)

Látens változó modell

- Vezessünk be látens változót:
 $z \sim p(z) = \mathcal{N}(0, I)$ (Prior)




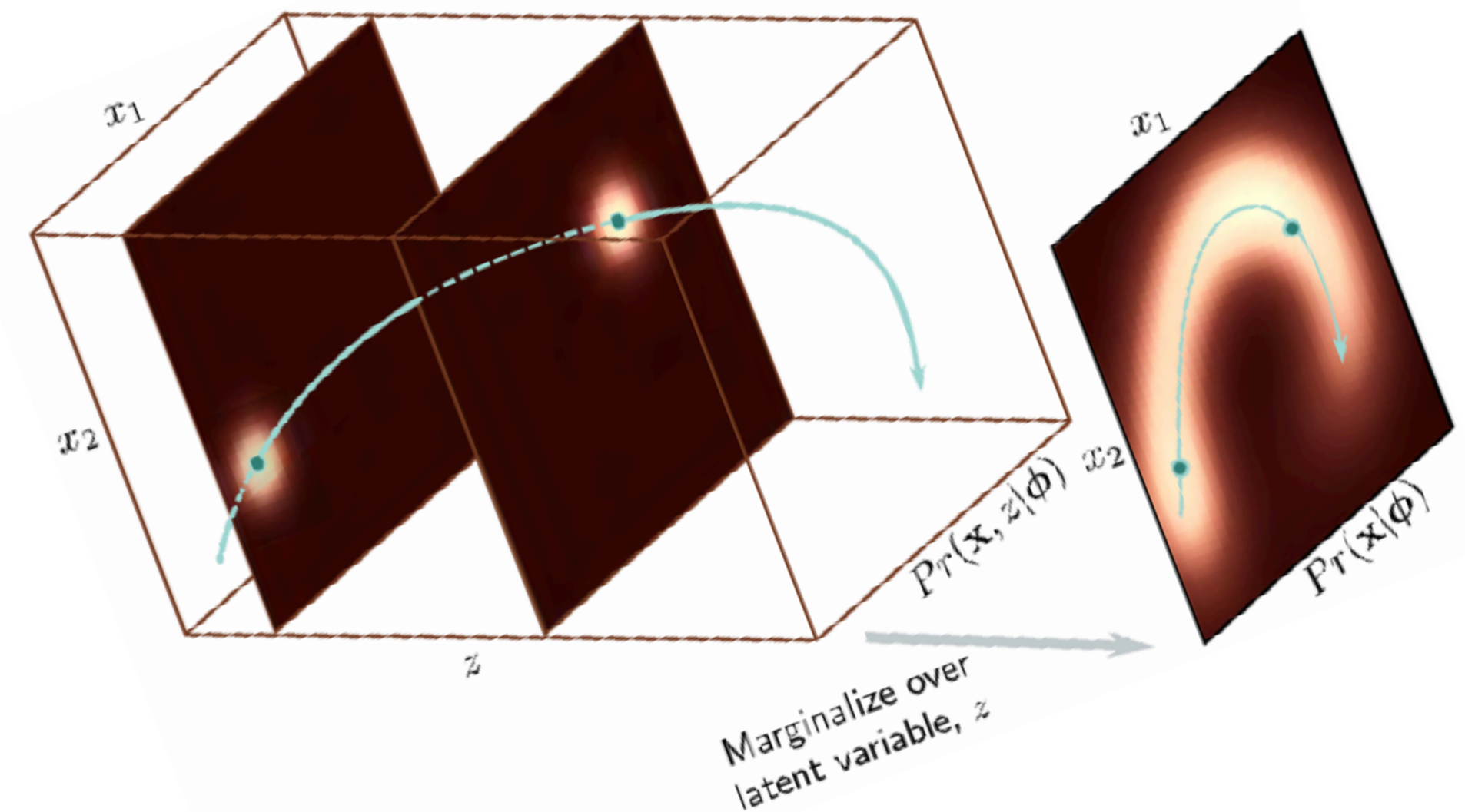
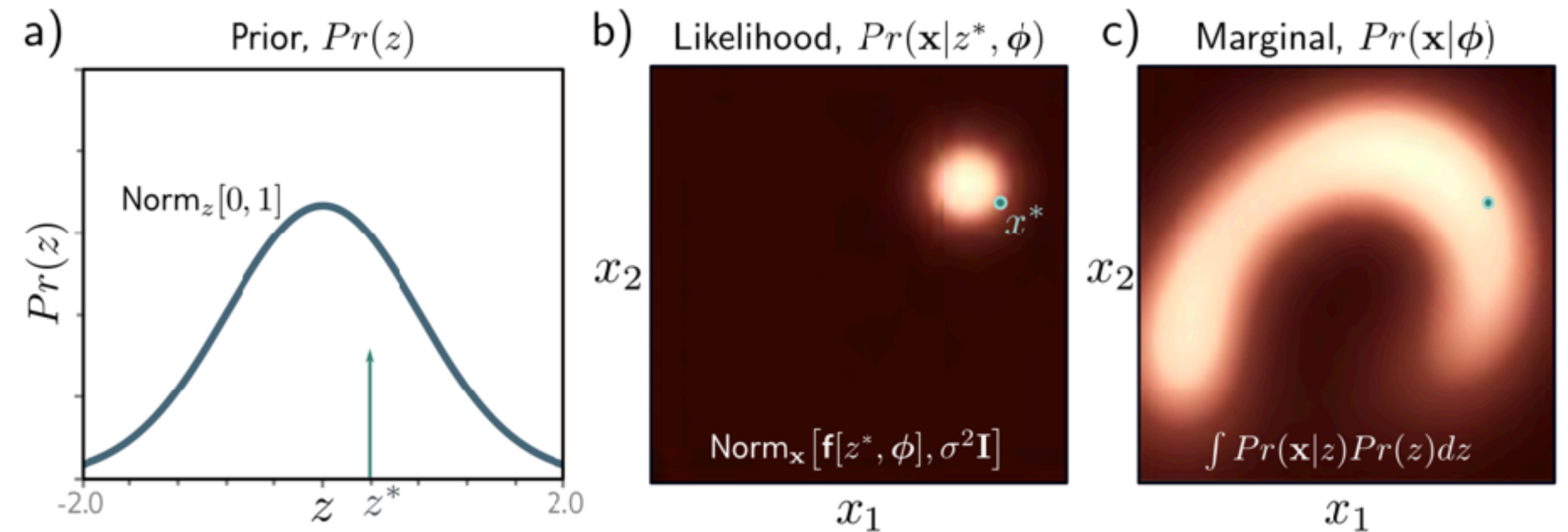
- Probabilisztikus dekóder:
 $p(x | z) \sim \mathcal{N}(\mu_{x|z}, \sigma)$

- Log-likelihood = MSE rekonstrukció:

$$\log p(x) = \log(e^{-\|x - \mu(z)\|^2}) = -\|x - \mu(z)\|^2 + \dots$$

- Likelihood (marginalizálás):

$$\log p(x) = \log \int p(x | z) \cdot p(z) dz$$




Variációs Autoenkóder (VAE)

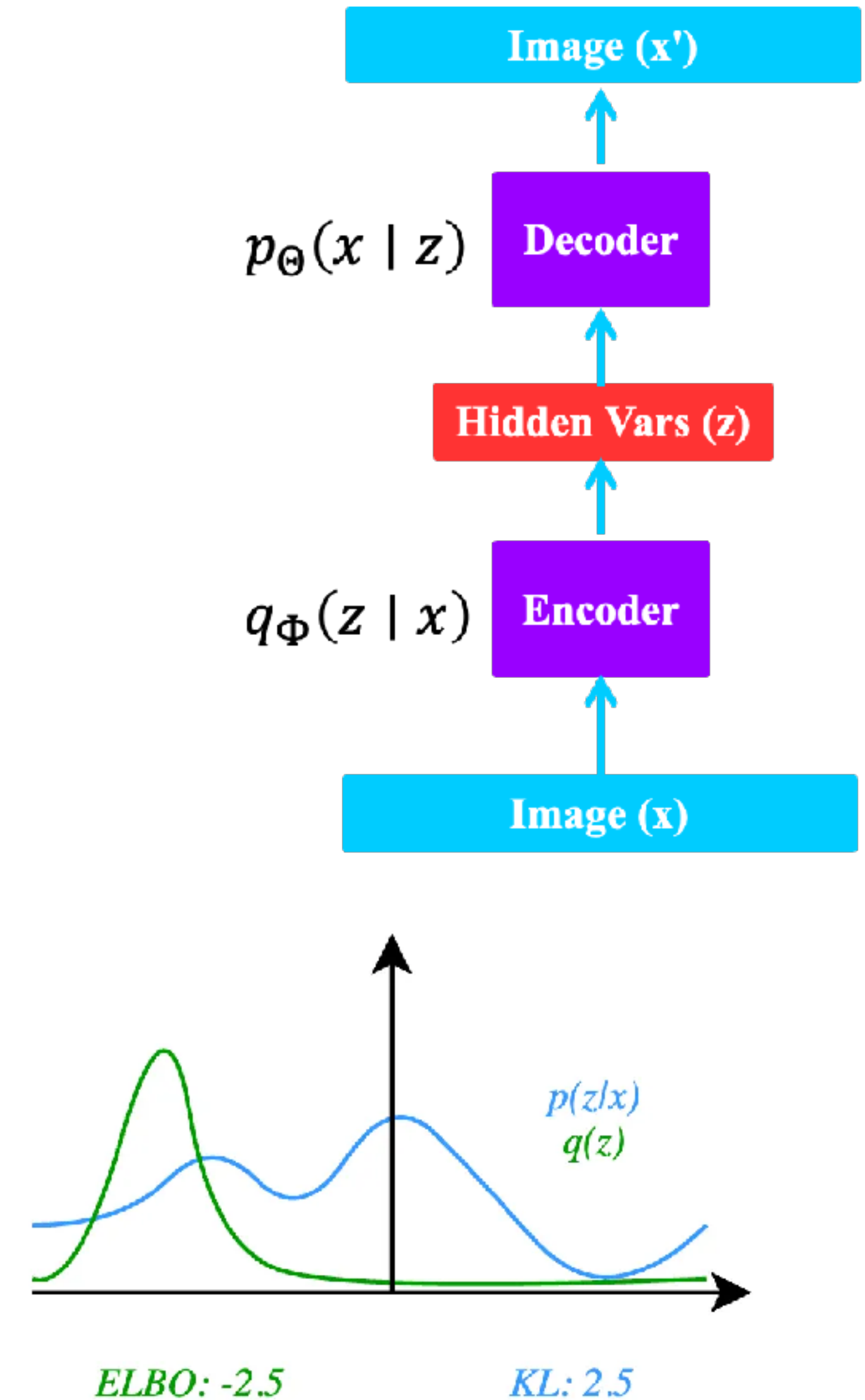
Variációs Bayes inferencia

- Cél: $dekoder = enkoder^{-1}$
- Probabilisztikus enkóder $p(z | x)$ — Bayes formula:

$$p(z | x) = \frac{\text{Dekóder} \quad \text{Prior}}{\text{Enkóder} \quad \text{(Posterior)}} = \frac{p(x | z) \cdot p(z)}{p(x)} \quad \text{☠}$$

- Variációs Bayes — közelítő posterior eloszlás
 $q(z | x) \sim \mathcal{N}(\mu_{z|x}, \Sigma_{z|x})$, minden bemenetre külön optimalizált paraméterekkel:

$$p(z | x) \approx \underset{q}{\operatorname{argmin}} \operatorname{KL}(q(z | x) || p(z | x))$$



Variációs Autoenkóder (VAE)

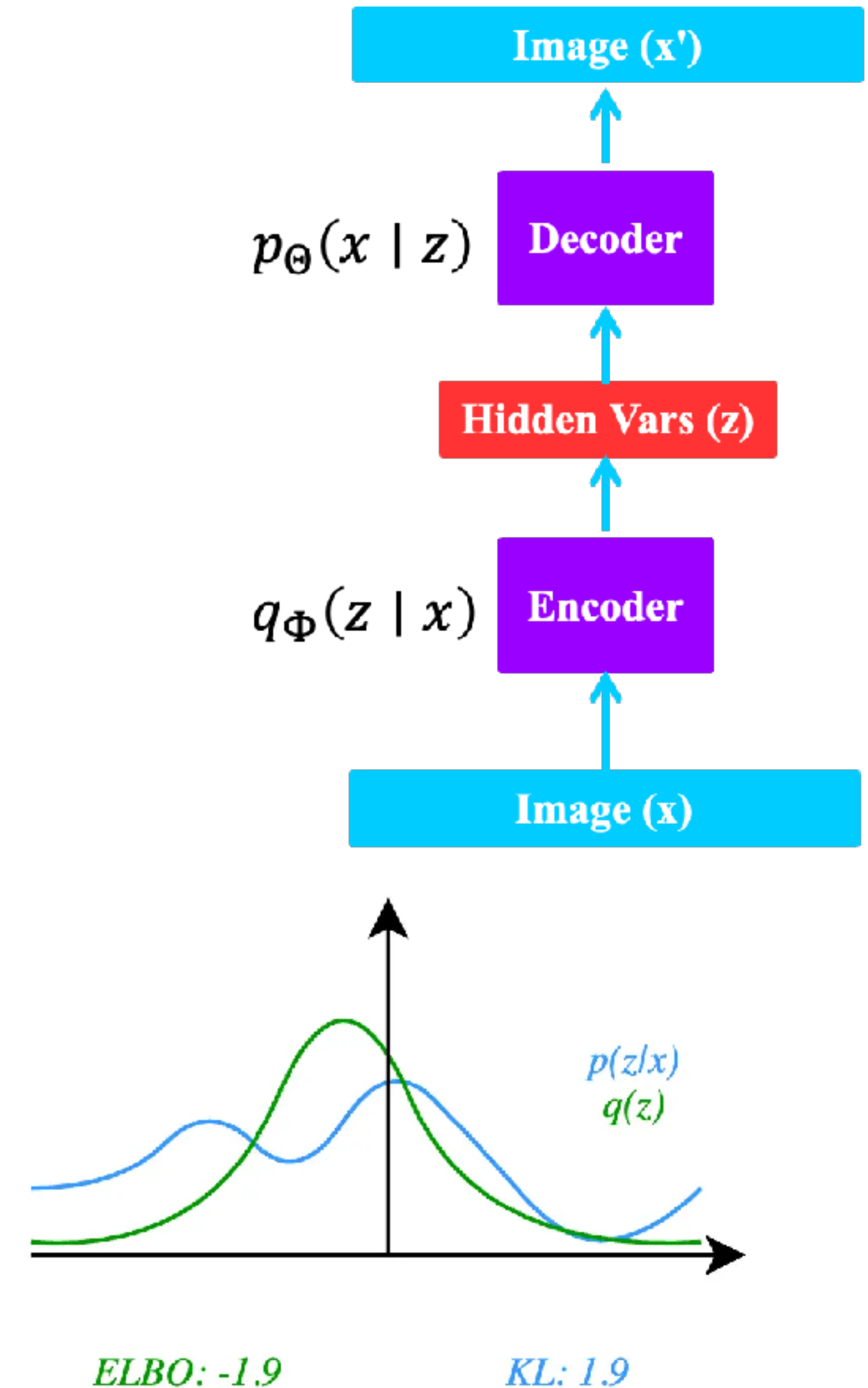
Variációs Bayes inferencia

- Cél: $dekoder = enkoder^{-1}$
- Probabilisztikus enkóder $p(z | x)$ — Bayes formula:

$$p(z | x) = \frac{\text{Dekóder} \quad \text{Prior}}{\text{Enkóder} \quad \text{(Posterior)}} = \frac{p(x | z) \cdot p(z)}{p(x)} \quad \text{☠}$$

- Variációs Bayes — közelítő posterior eloszlás
 $q(z | x) \sim \mathcal{N}(\mu_{z|x}, \Sigma_{z|x})$, minden bemenetre külön optimalizált paraméterekkel:

$$p(z | x) \approx \underset{q}{\operatorname{argmin}} \operatorname{KL}(q(z | x) || p(z | x))$$



Variációs Autoenkóder (VAE)

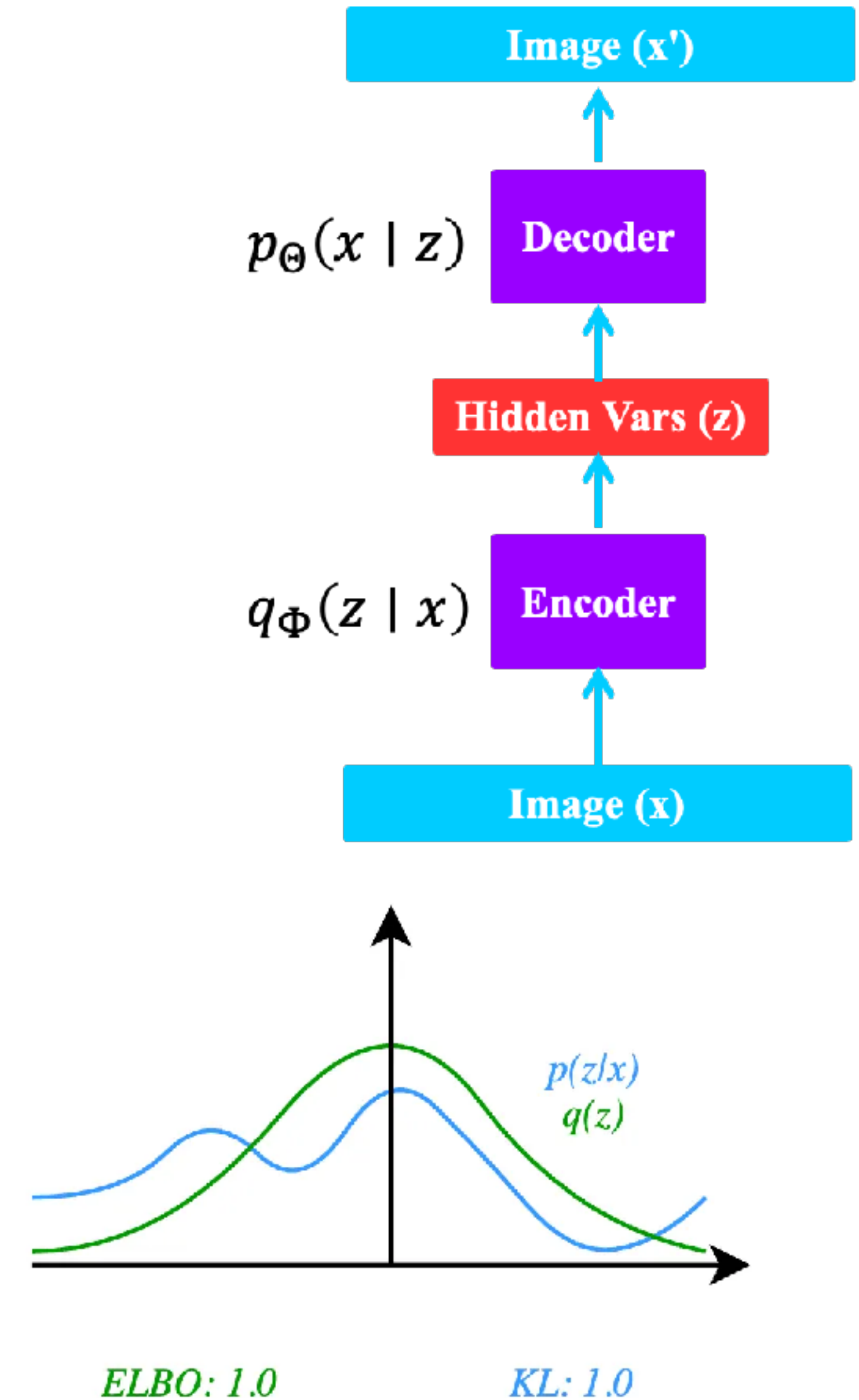
Variációs Bayes inferencia

- Cél: $dekoder = enkoder^{-1}$
- Probabilisztikus enkóder $p(z | x)$ — Bayes formula:

$$p(z | x) = \frac{\text{Dekóder} \quad \text{Prior}}{\text{Enkóder} \quad \text{(Posterior)}} = \frac{p(x | z) \cdot p(z)}{p(x)} \quad \text{☠}$$

- Variációs Bayes — közelítő posterior eloszlás
 $q(z | x) \sim \mathcal{N}(\mu_{z|x}, \Sigma_{z|x})$, minden bemenetre külön optimalizált paraméterekkel:

$$p(z | x) \approx \underset{q}{\operatorname{argmin}} \operatorname{KL}(q(z | x) || p(z | x))$$



Variációs Autoenkóder (VAE)

Evidence Lower Bound (ELBO)

A végső célunk: $\max \log p(x)$

- Egy kis favágás...

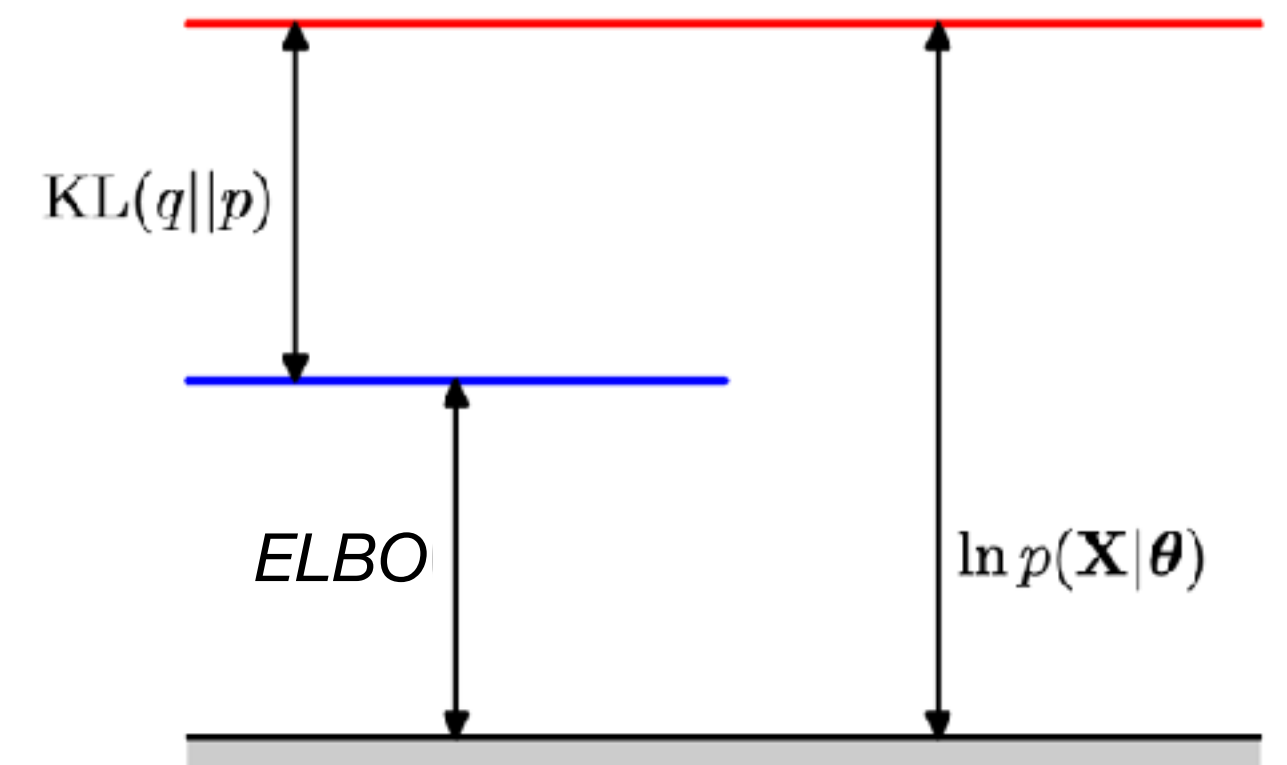
$$\begin{aligned} \text{KL}(q(z|x)||p(z|x)) &= \mathbb{E}_{q(z|x)} [\log q(z|x)] - \mathbb{E}_{q(z|x)} [\log p(z|x)] \\ &= \mathbb{E}_{q(z|x)} [\log q(z|x)] - \mathbb{E}_{q(z|x)} [\log p(x|z)] - \mathbb{E}_{q(z|x)} [\log p(z)] + \mathbb{E}_{q(z|x)} [\log p(x)] \\ &= - \underbrace{(\mathbb{E}_{q(z|x)} [\log p(x|z)] - \text{KL}(q(z|x)||p(z)))}_{ELBO} + \log p(x) \end{aligned}$$

- Azaz:

$$\log p(x) = ELBO + \underbrace{\text{KL}(q(z|x)||p(z|x))}_{\geq 0!!}$$

“model evidence”

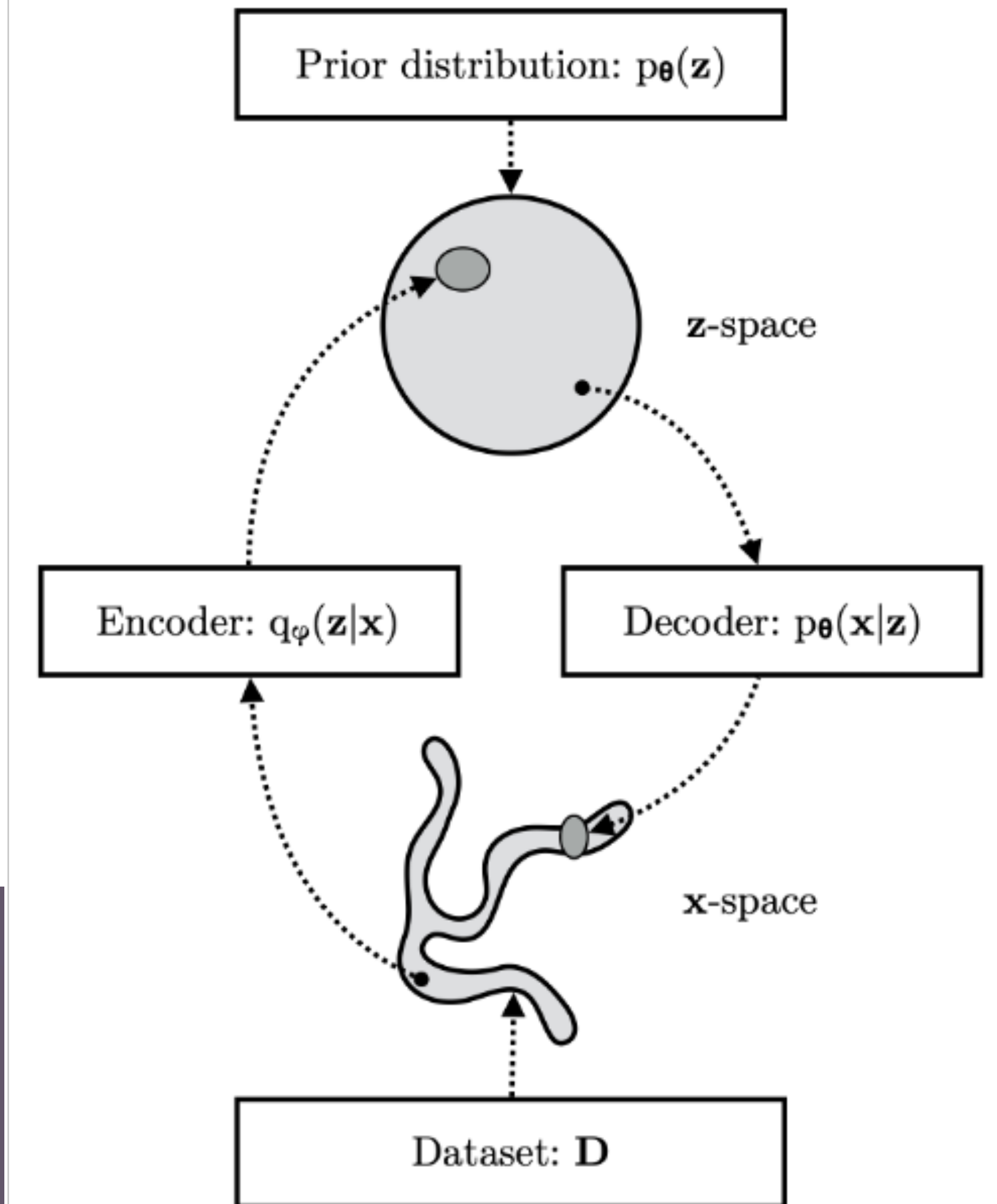
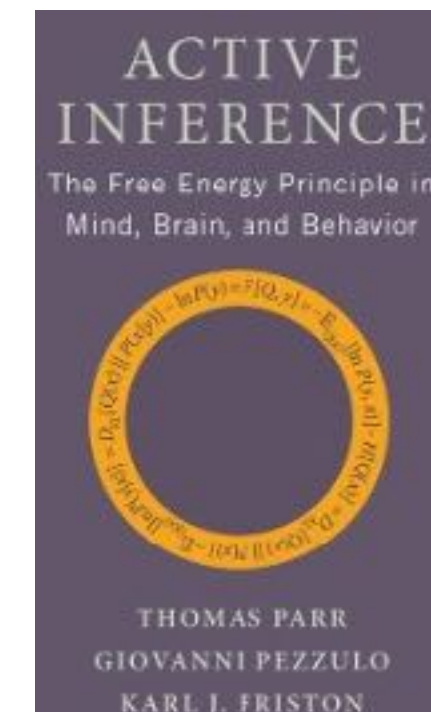
- $\log p(x) \geq ELBO \Rightarrow$ **Evidence Lower Bound**



Variációs Autoenkóder (VAE)

ELBO – Ekvivalens alakok

- $ELBO = \log p(x) - \text{KL}(q(z | x) || p(z | x))$
- $ELBO = \mathbb{E}_{q(z|x)}[\log p(x | z)] + \text{KL}(q(z | x) || p(z))$
rekonstrukciós hiba + regularizáció
- $ELBO = \mathbb{E}_{q(z|x)}[\log p(x, z)] - H(q(z | x))$
teljes energia – entrópia
(variációs) szabad energia



Variációs Autoenkóder (VAE)

VAE loss

- Variációs Autoenkóder (VAE) – ELBO loss (1 db x^i adatpontra):

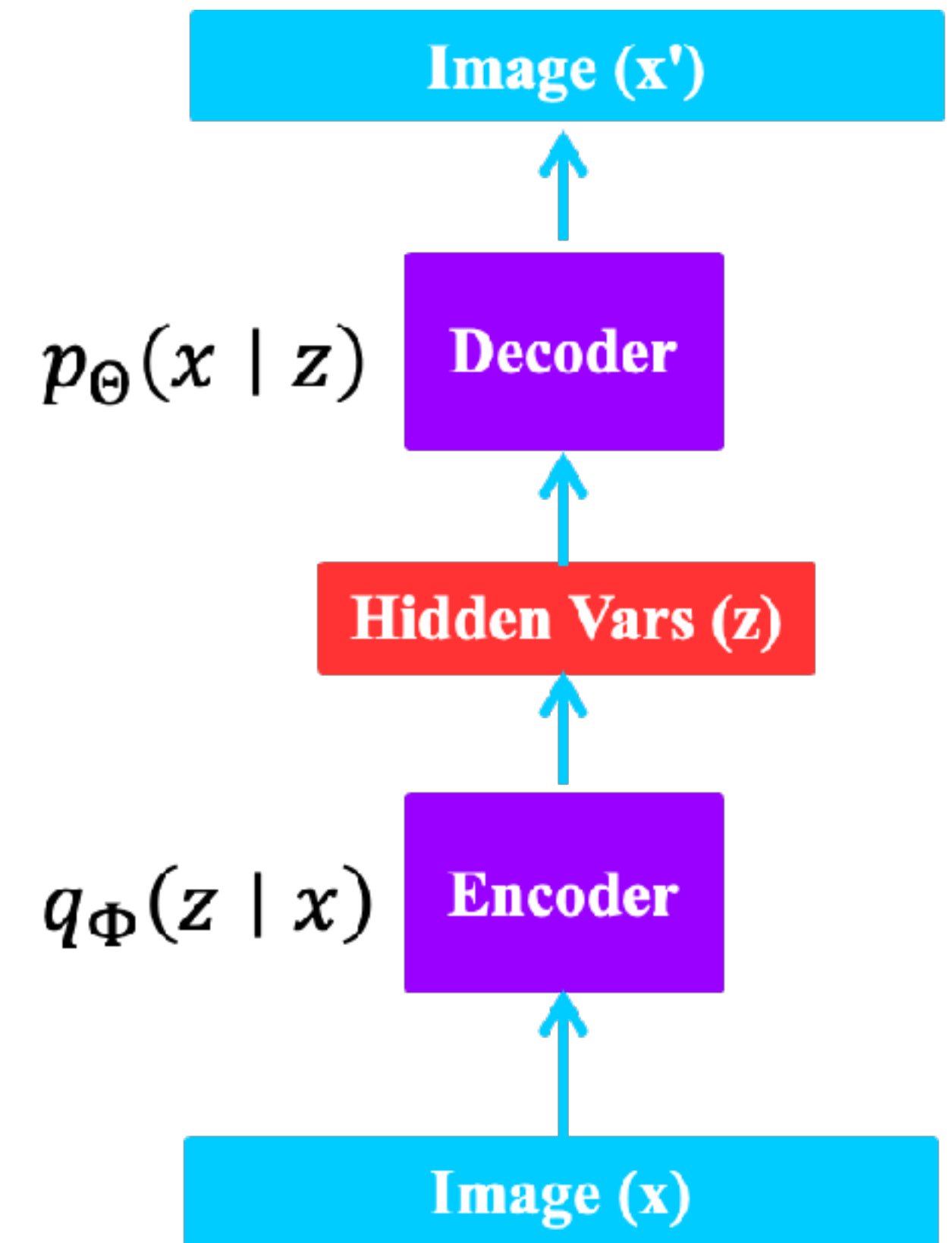
$$L_{VAE}(x^i, \Theta, \Phi) = \underbrace{\mathbb{E}_{q_{\Phi}(z|x^i)}[\log p_{\Theta}(x^i | z)]}_{\text{rekonstrukciós hiba}} - \underbrace{\text{KL}(q_{\Phi}(z | x^i) || p(z))}_{\text{regularizáció}}$$

- Gauss eloszlásokra ($q_{\Phi}(z | x^i) = \mathcal{N}(\mu_{\Phi}, \text{diag}(\sigma_{\Phi}))$):

$$L_{VAE}(x^i, \Theta, \Phi) = \mathbb{E}_{z \sim q_{\Phi}(z|x^i)}[\|x - D_{\Theta}(z)\|^2] - \sum_{i=1}^{|z|} (\mu_{i,\Phi}^2(x) + \sigma_{i,\Phi}^2(x) - 2 \ln(\sigma_{i,\Phi}))$$

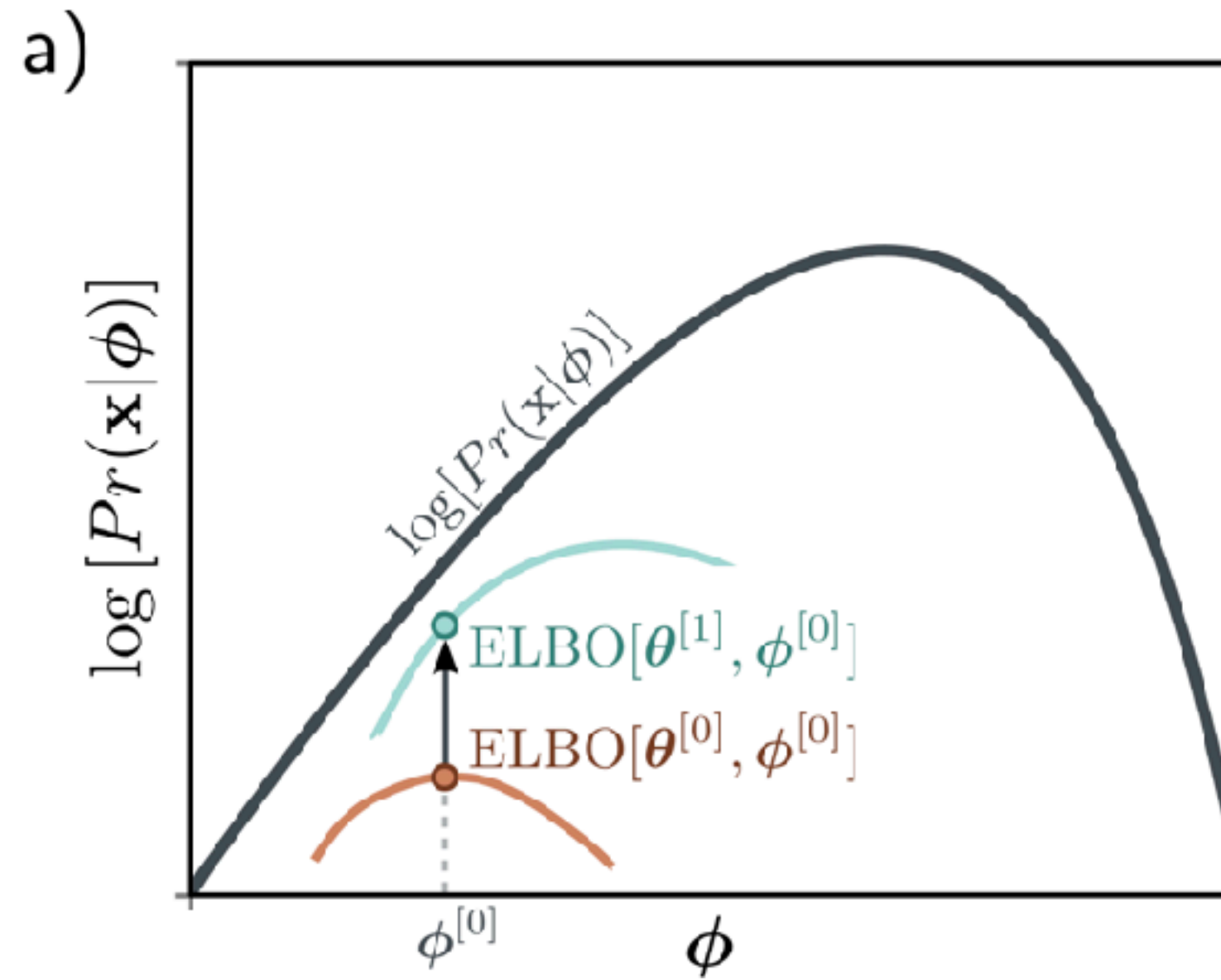
- Több adatra (mini-batch-re) átlagoljuk:

$$\mathbb{E}_x[L_{VAE}(x, \Theta, \Phi)] \rightarrow \min$$

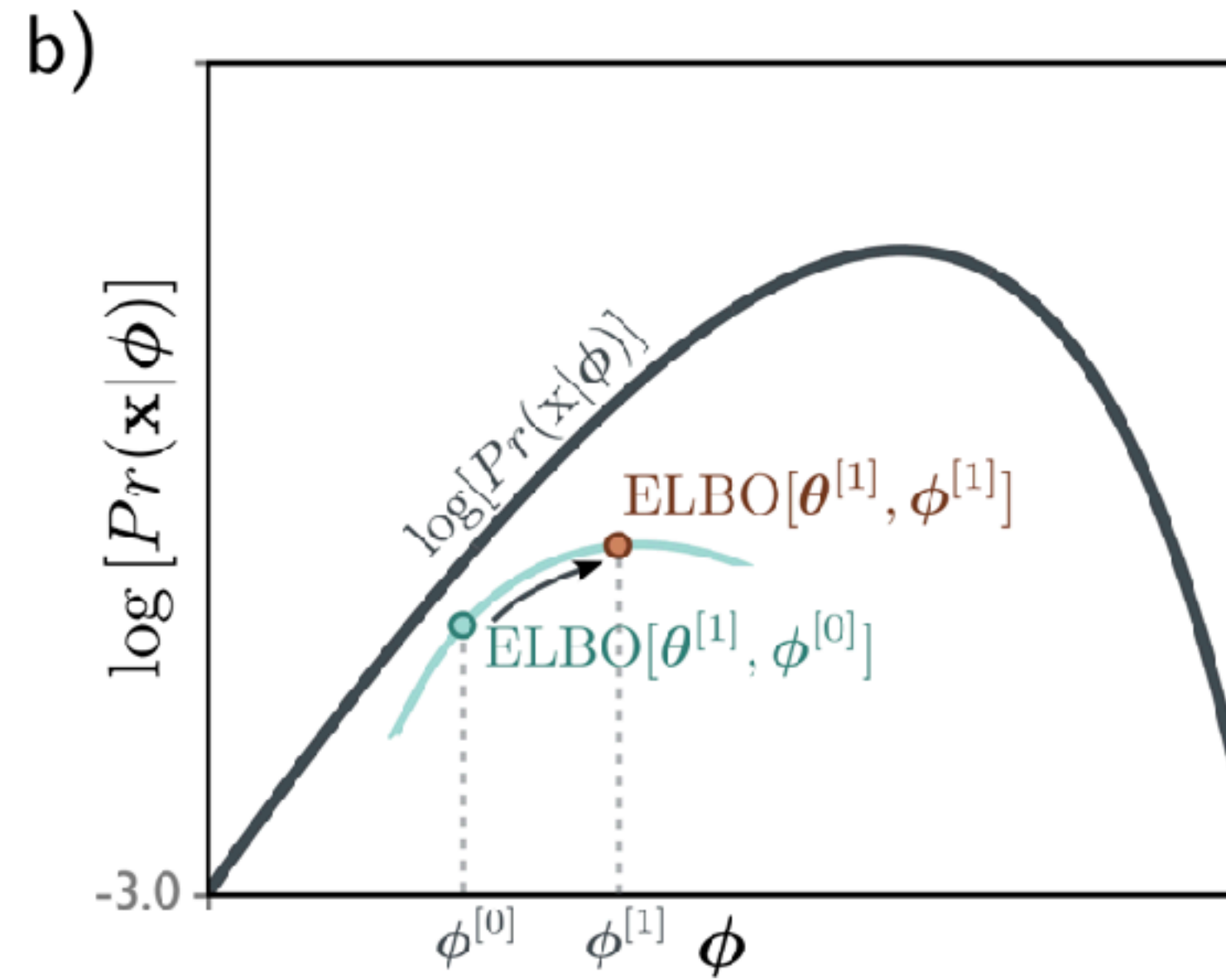


Variációs Autoenkóder (VAE)

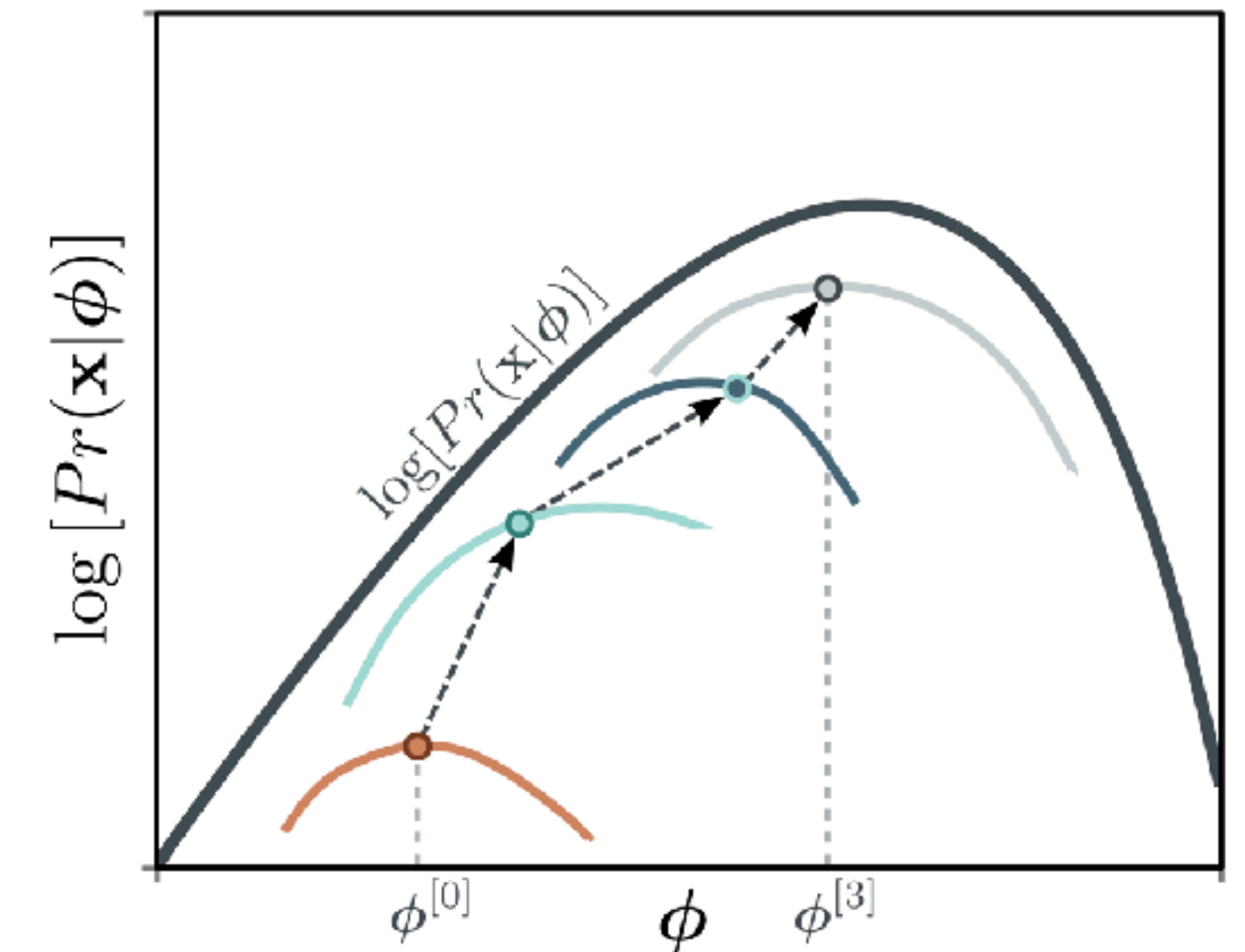
Tanítás



Dekóder optimalizálás



Enkóder optimalizálás

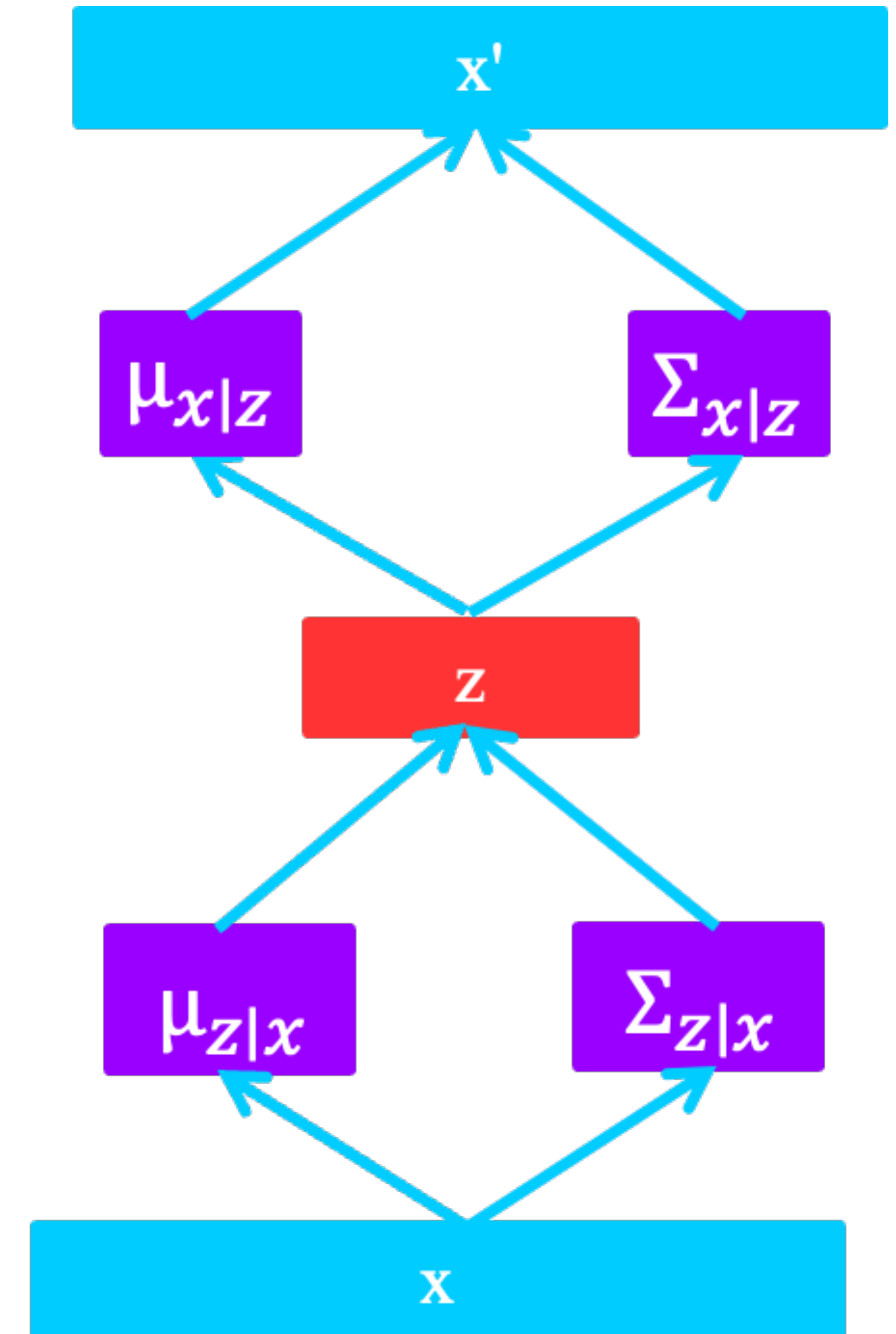


End-to-end optimalizálás

Variációs Autoenkóder (VAE)

Tanítás – Amortizált inferencia

- Minden x -re külön optimalizáljuk a közelítő posterior $\mu_{z|x}$, $\Sigma_{z|x}$ paramétereit?
- **Amortizált inferencia**: az eloszlások paramétereit állítsa elő egy neurális háló!
- VAE tanítása
Minden x minibatch-re:
 1. Mintavétel (enkóder): $z \sim \mathcal{N}(\mu_{z|x}, \Sigma_{z|x})$
 2. Mintavétel (dekóder): $x' \sim \mathcal{N}(\mu_{x|z}, \Sigma_{x|z})$
 3. Loss kiértékelés: $\mathbb{E}_x[L_{VAE}(x, \Theta, \Phi)]$
 4. Θ, Φ backprop + optimalizáció



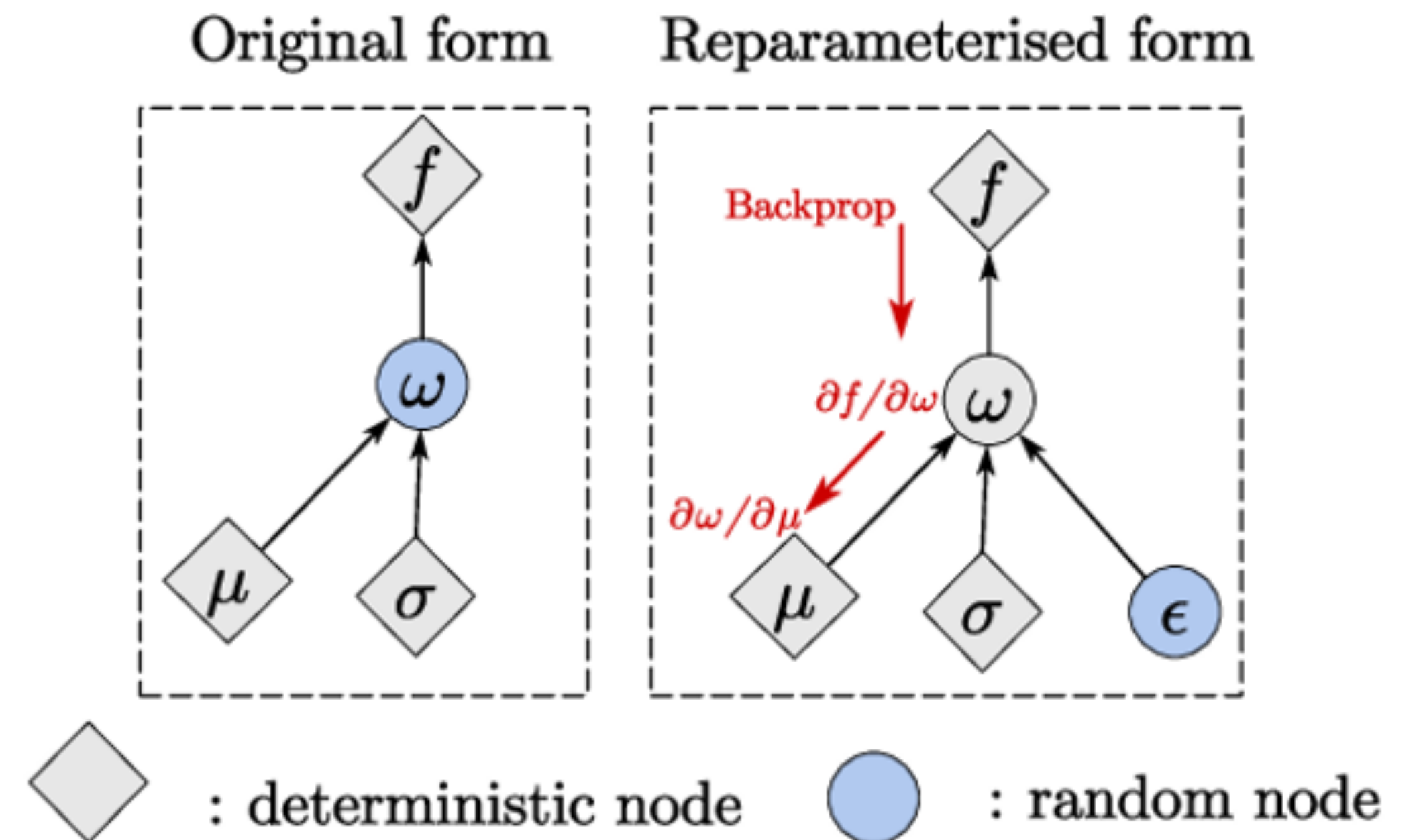
Variációs Autoenkóder (VAE)

Tanítás, inferencia – Reparametrizációs trükk

- Hogyan optimalizálunk egy $z \sim p(z | x) = \mathcal{N}(\mu, \Sigma)$ eloszlásból mintavételezett random változón keresztül?
- Deriváljuk a mintavételezés műveletét?
- A kiskapu – reparametrizációs trükk:

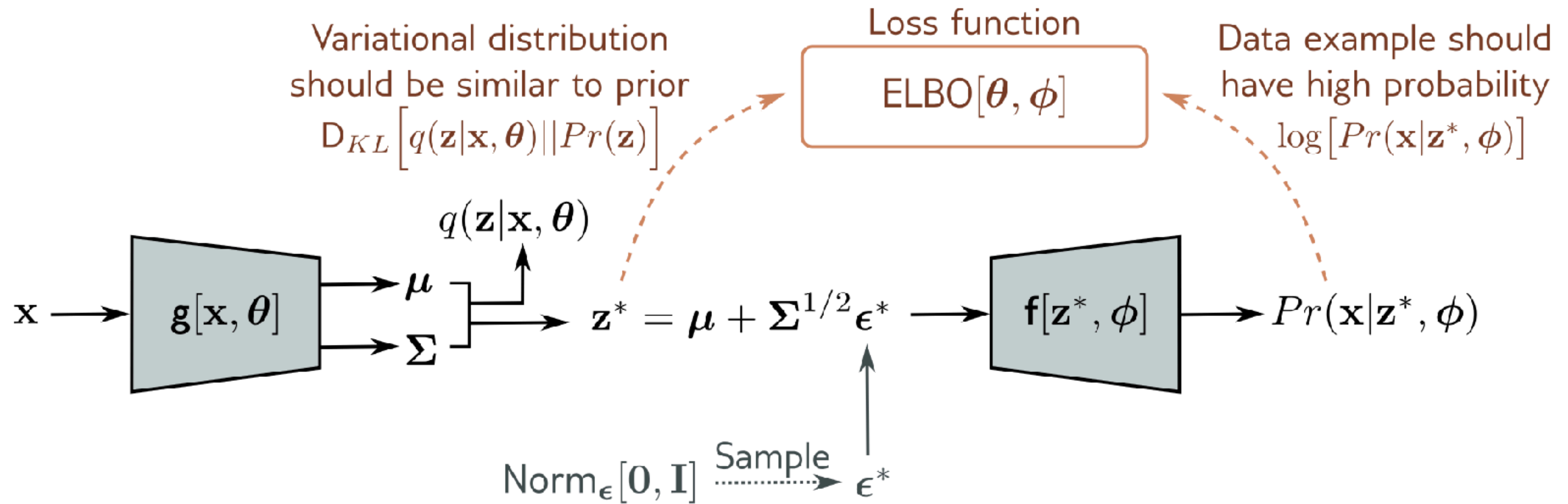
$$z = \mu + \sigma \cdot \epsilon, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- Az optimalizált változók szerint determinisztikus függvény!



Variációs Autoenkóder (VAE)

Összefoglalás



Variációs Autoenkóder (VAE)

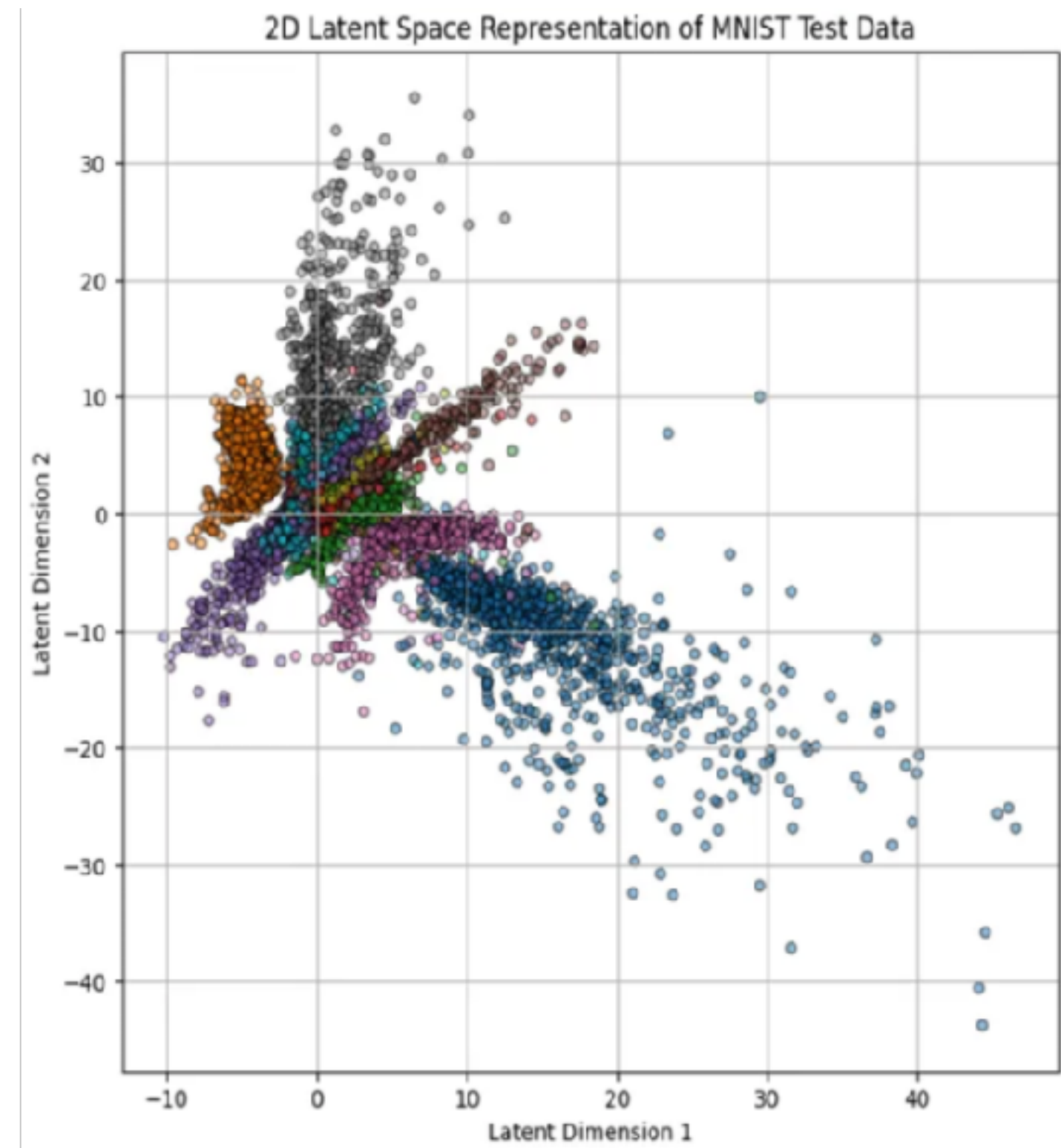
Egy tipikus állapot a VAE tanulása közben:



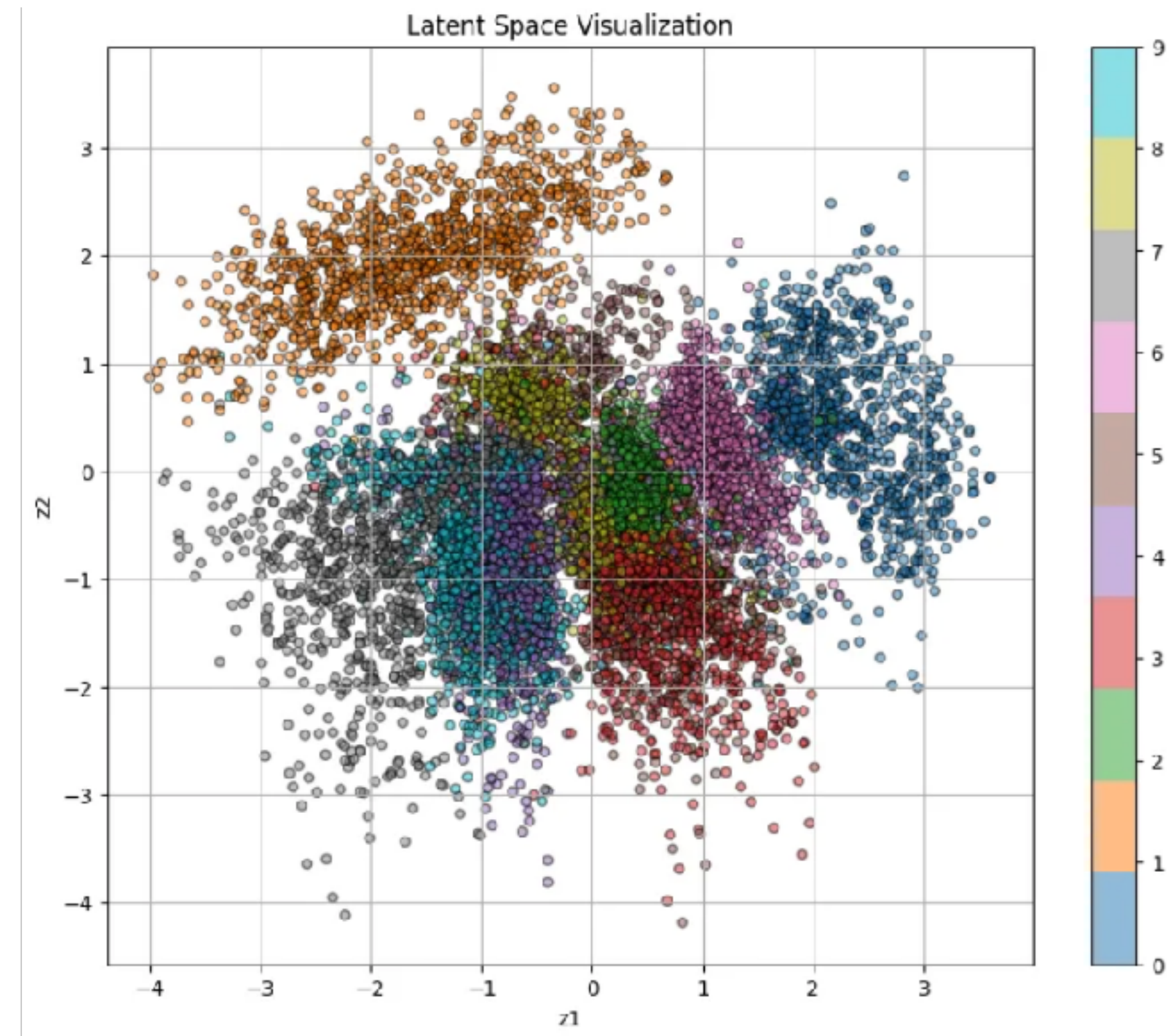
Elsőre (meg másodikra, sít.) nem könnyű átlátni, de ne adjuk fel!
Gondoljuk át a lépéseket alaposan és ha valami nem tiszta, kérdezzünk!

Variációs Autoenkóder (VAE)

AE vs. VAE – Látens tér



AE



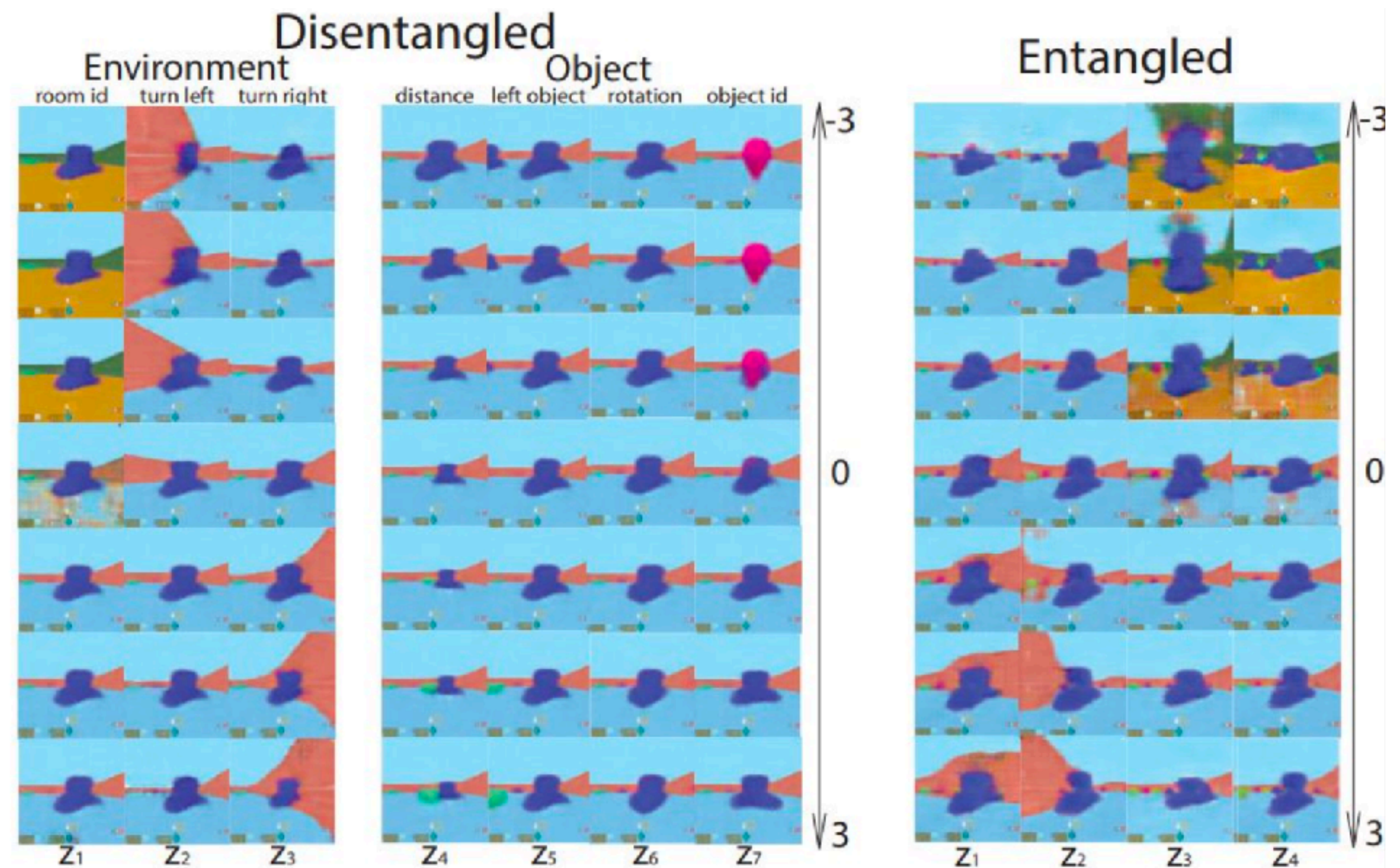
VAE

Variációs Autoenkóder (VAE)

Variációk – β -VAE

$$\mathbb{E}_{q_{\Phi}(z|x^i)}[\log p_{\Theta}(x^i | z)] - \beta \text{KL}(q_{\Phi}(z | x^i) || p(z))$$

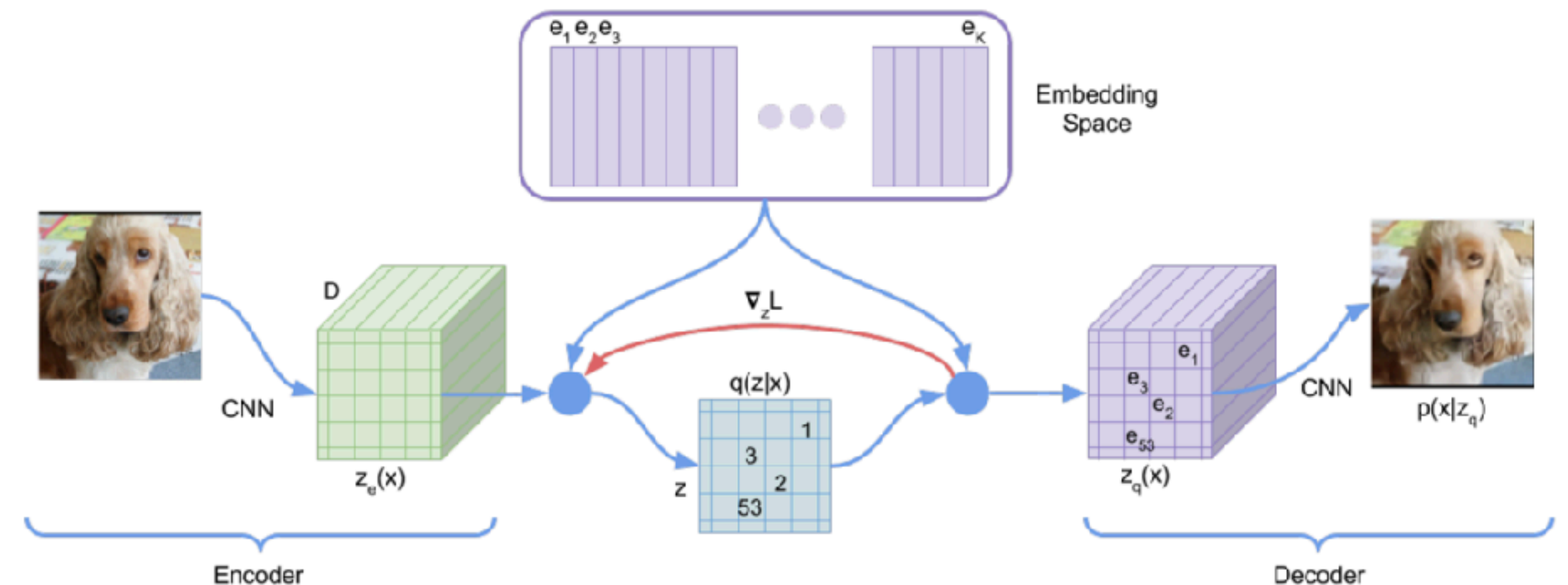
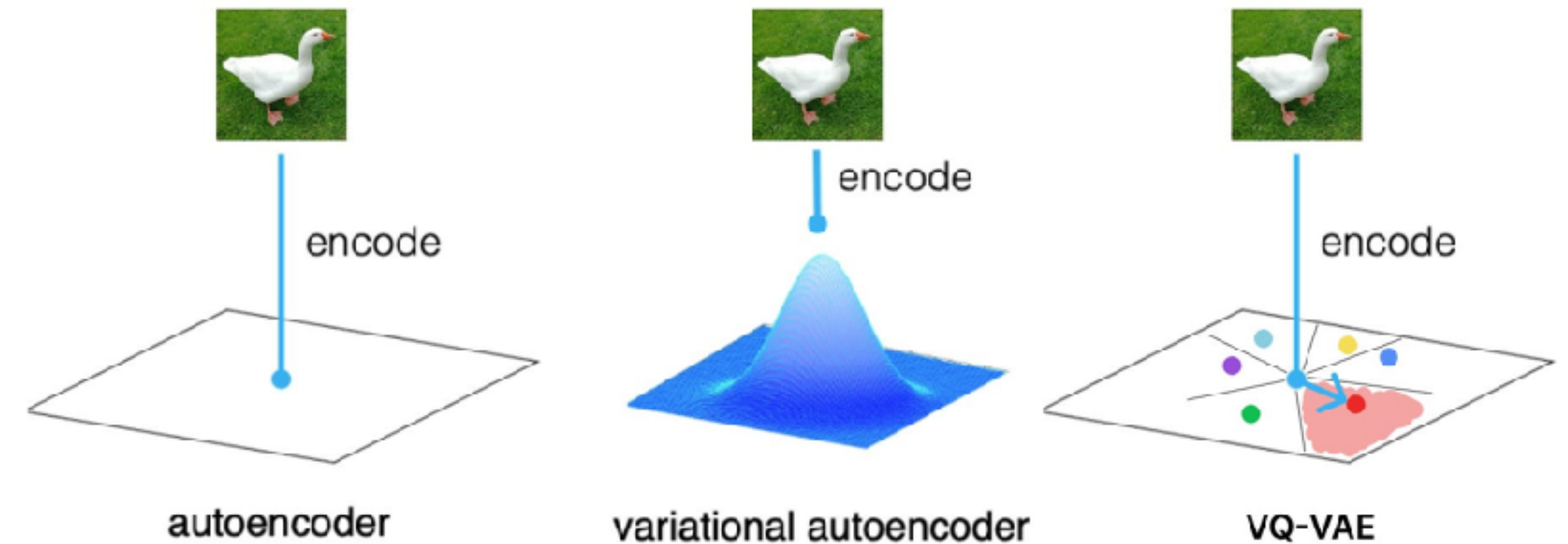
A regularizációs tagot súlyozva függetleníteni lehet a látens dimenziók hatását!



Variációs Autoenkóder (VAE)

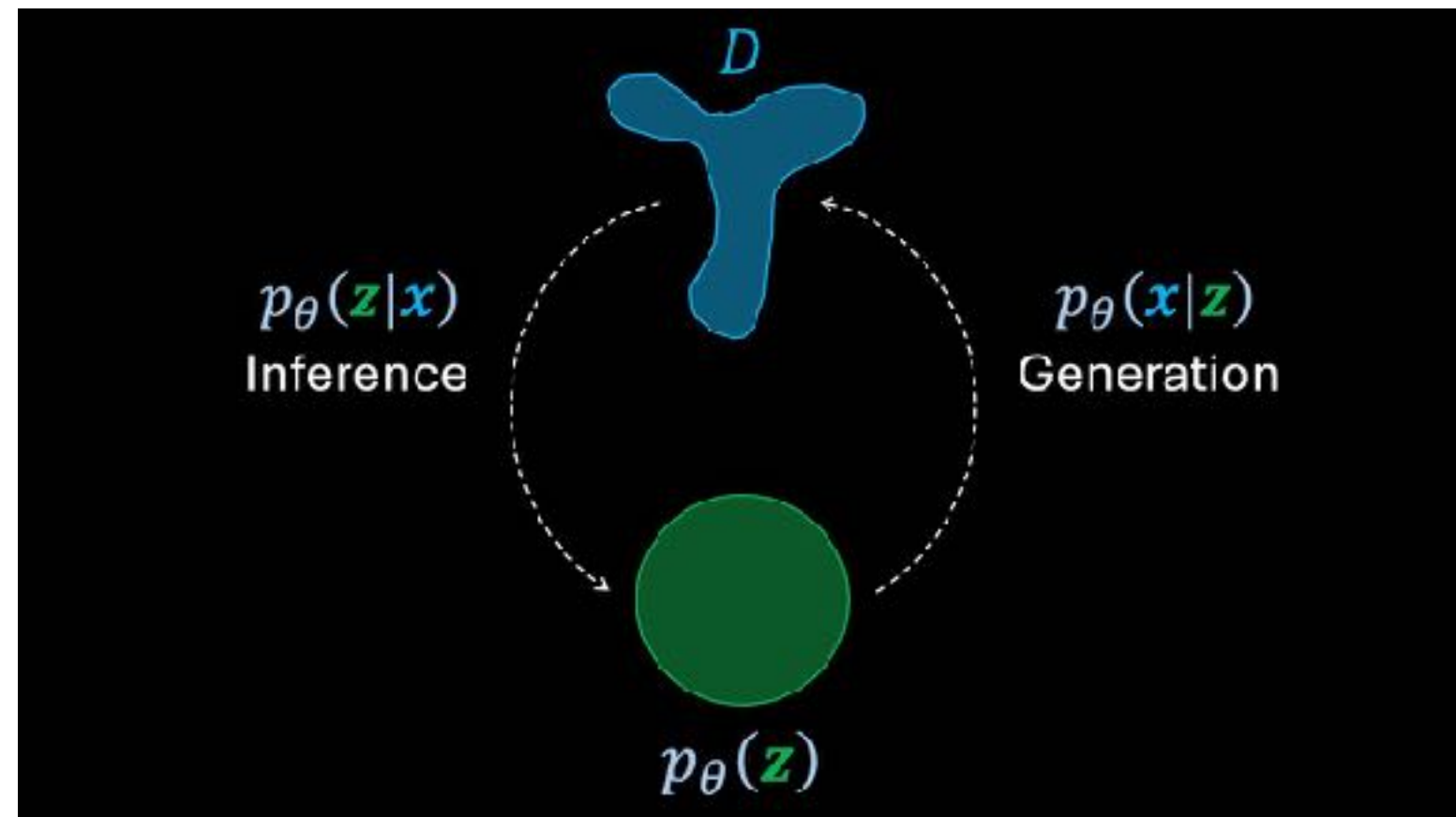
Variációk – VQVAE

- Ötlet: *kvantáljuk* a látens kódokat (diszkrét **codebook**)
- **Vektor kvantálás (VQ)**, pl. Nearest Neighbor / K-Means
- “**Tokenizálás**” – lehetővé teszi pl. az autoregresszív generálást!
- Tömöríthető reprezentáció
- Egyfajta tanítható prior

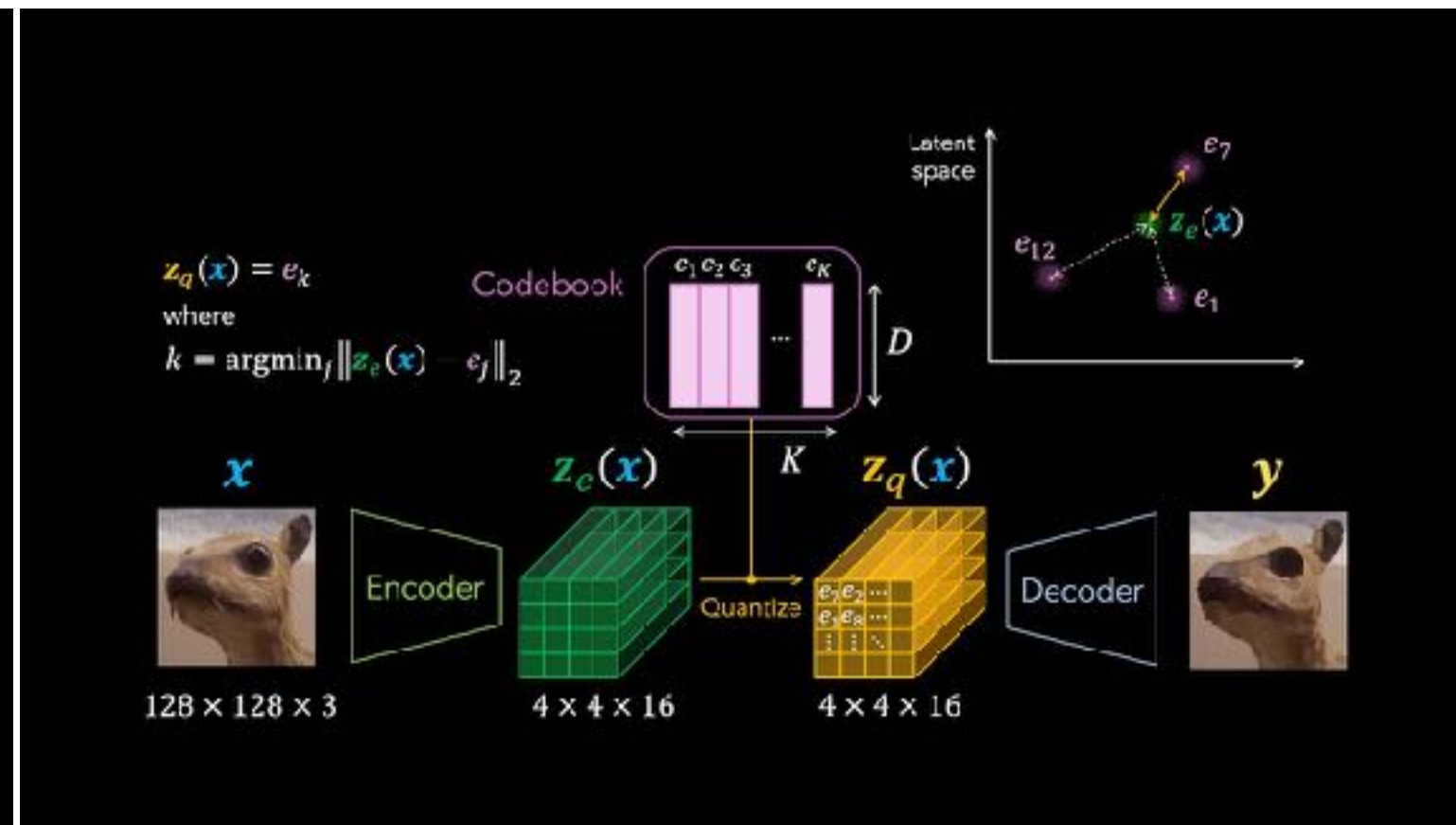


Variációs Autoenkóder (VAE)

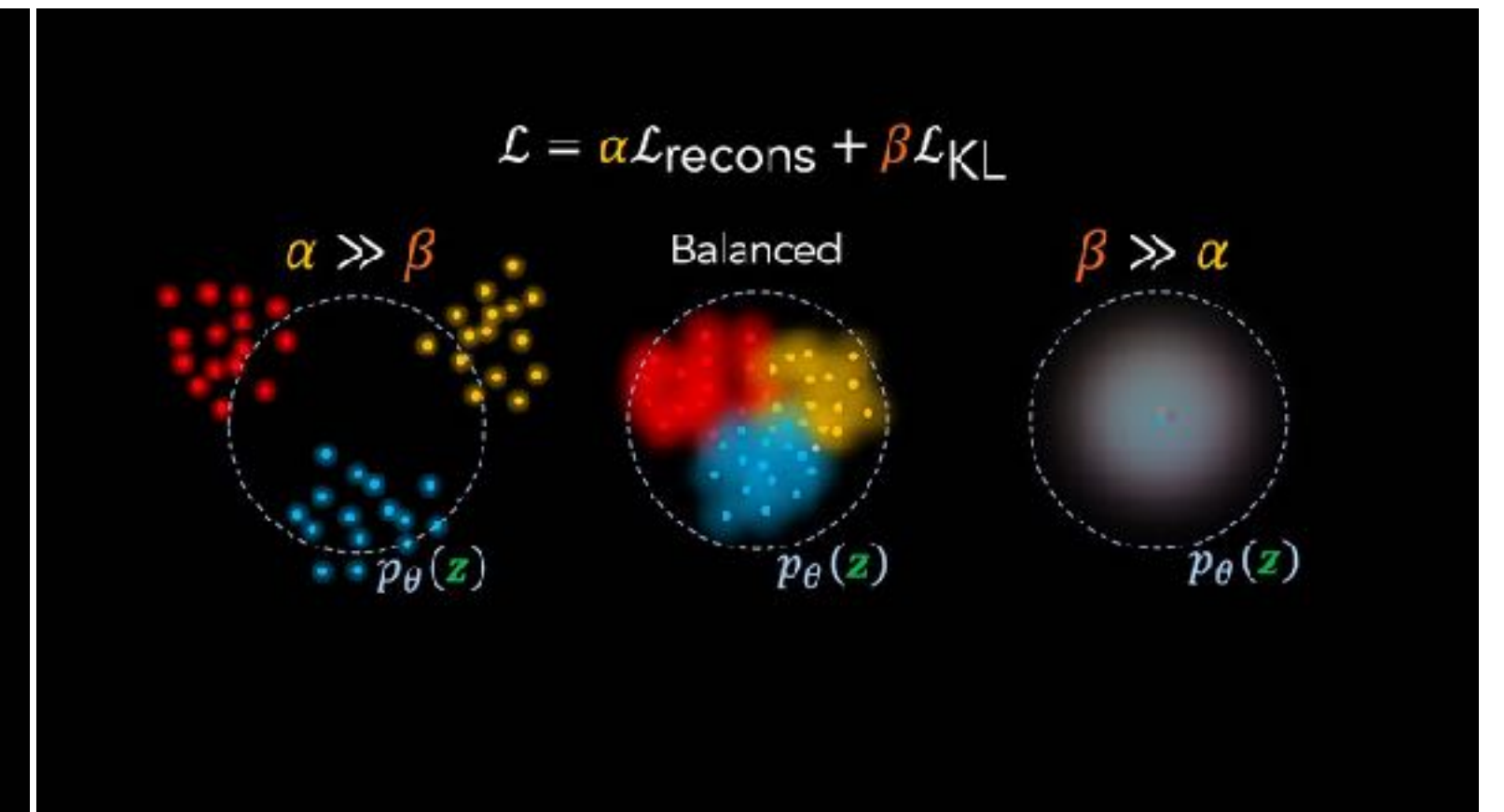
Videóajánló



VAE



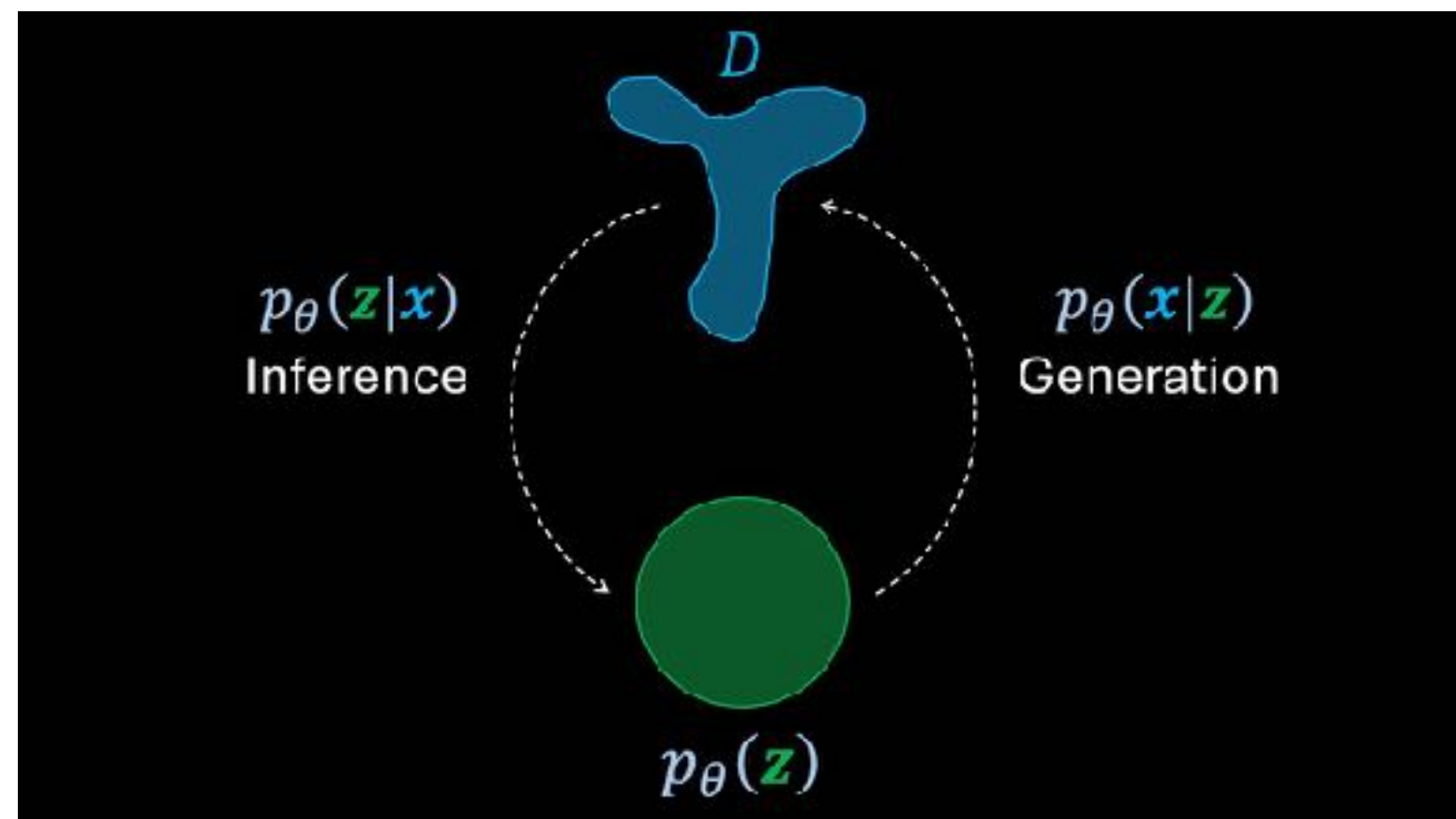
VQVAE



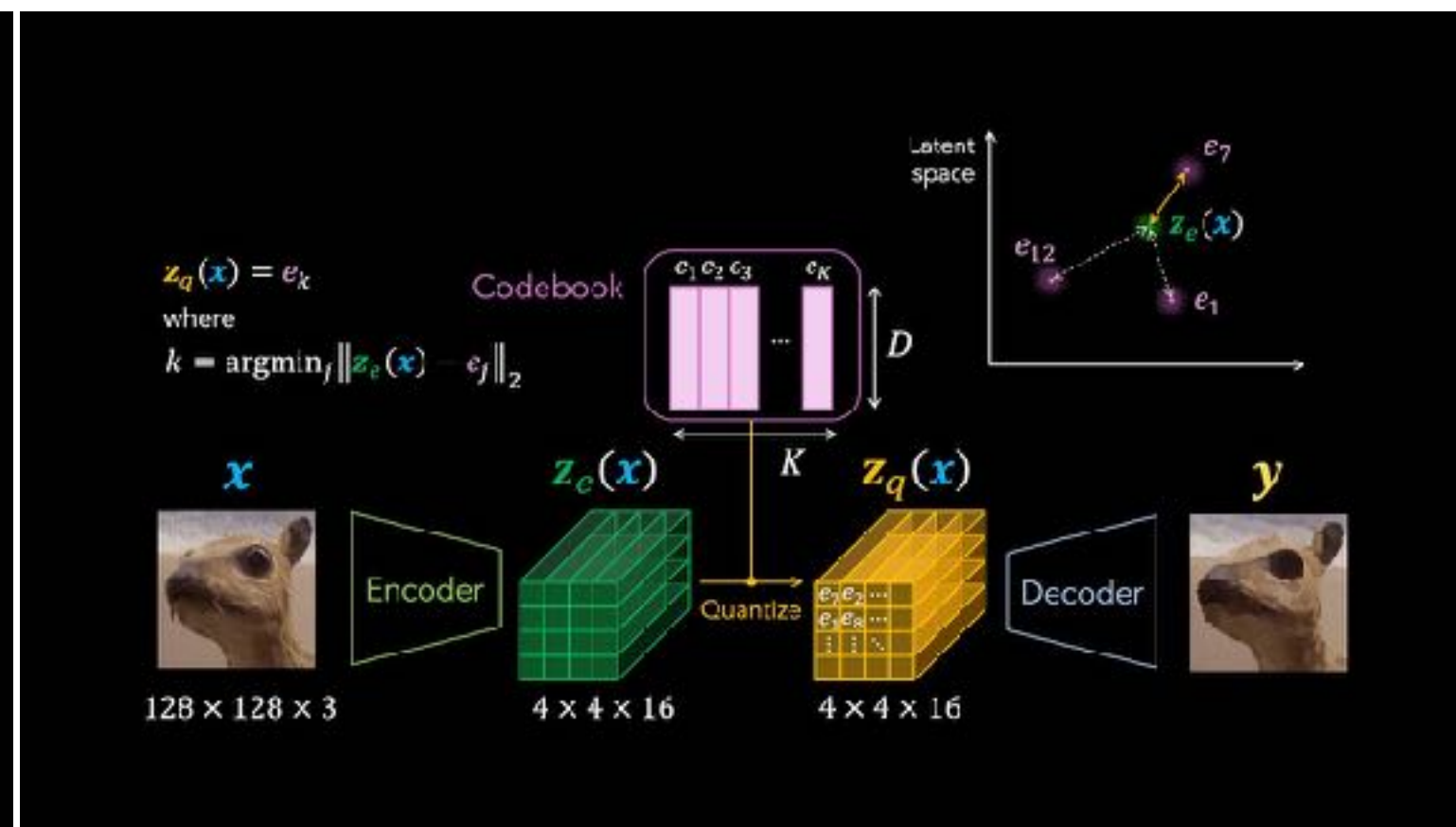
β -VAE

Variációs Autoenkóder (VAE)

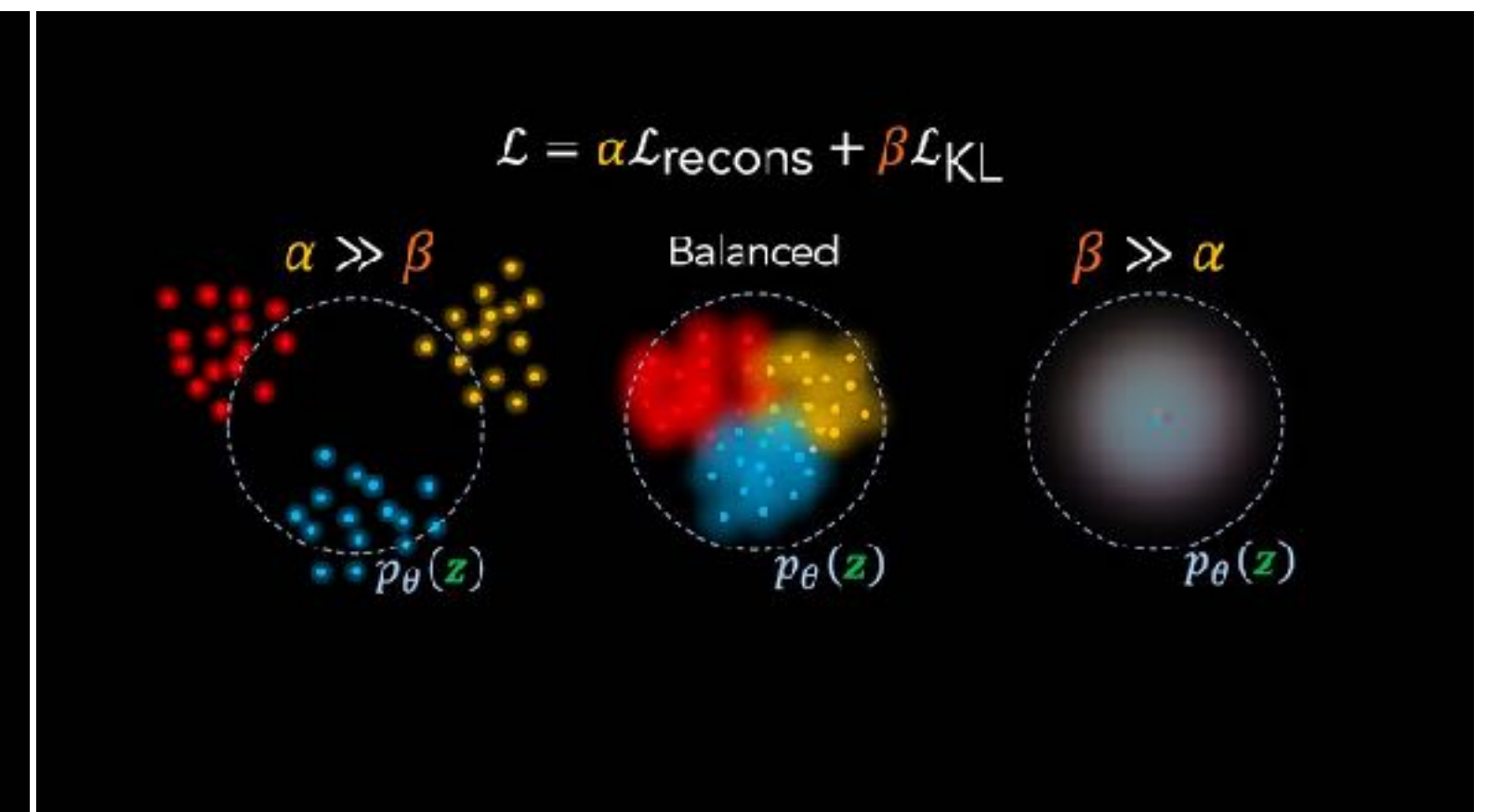
Videóajánló



VAE



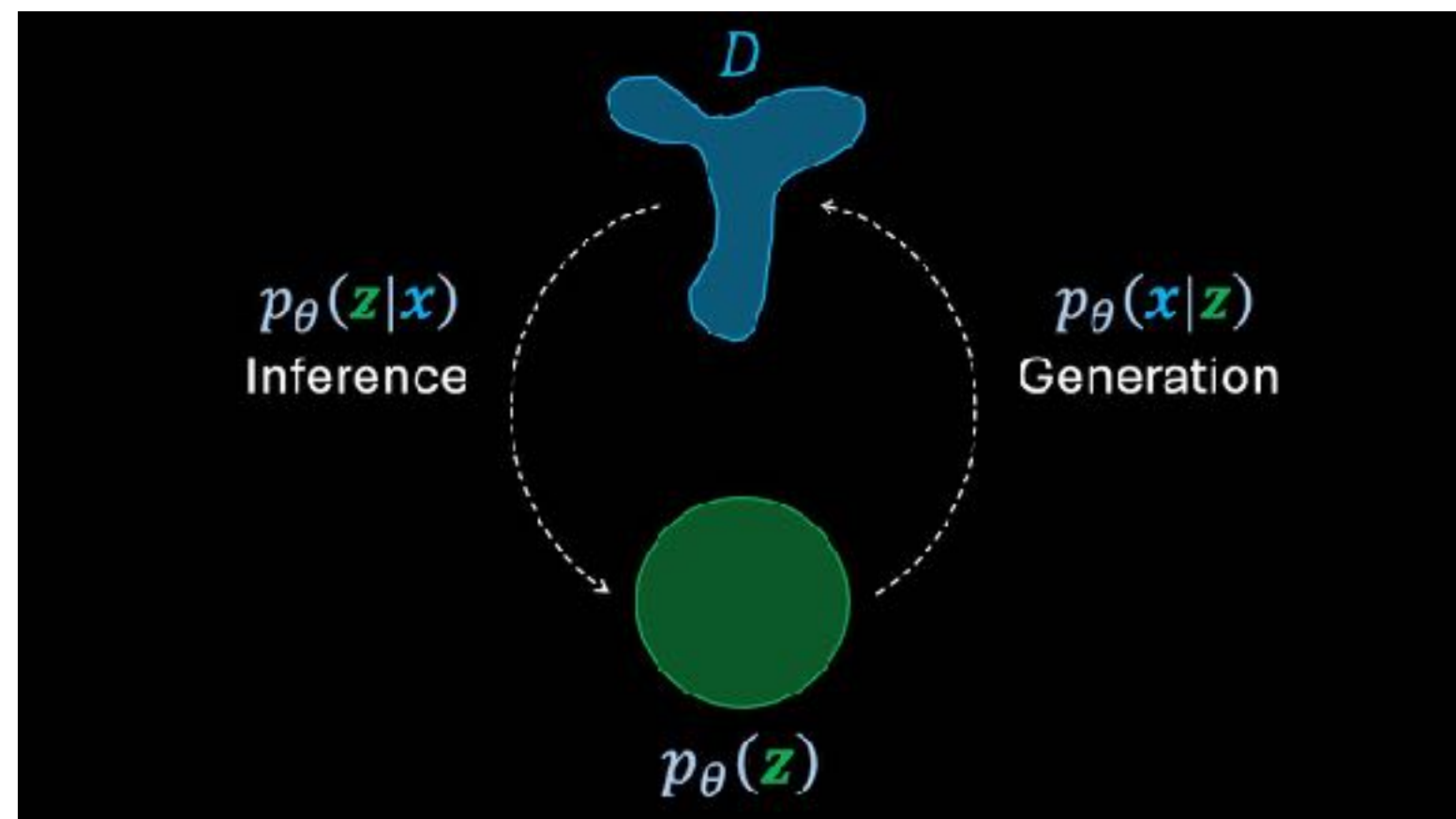
VQVAE



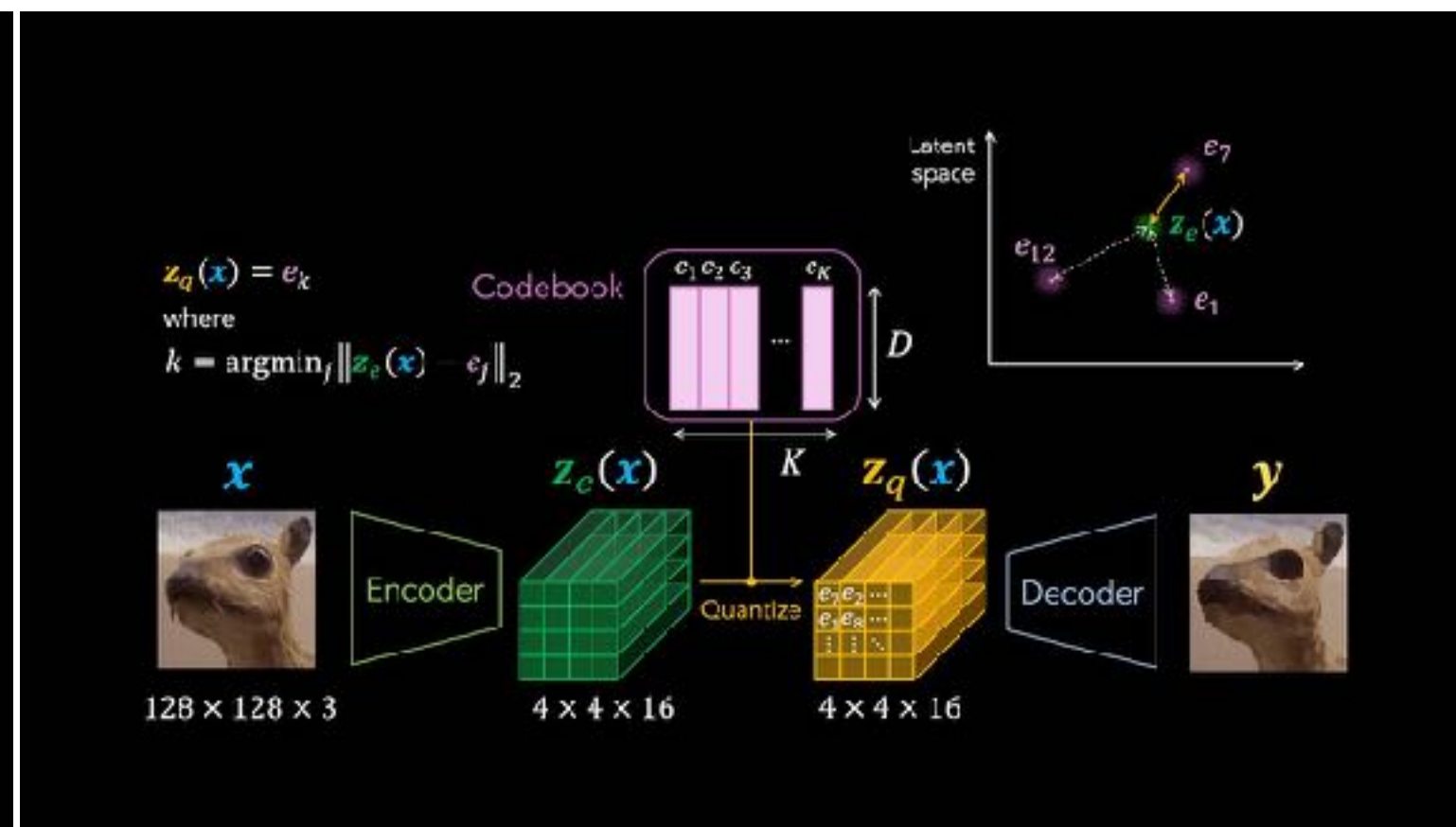
β -VAE

Variációs Autoenkóder (VAE)

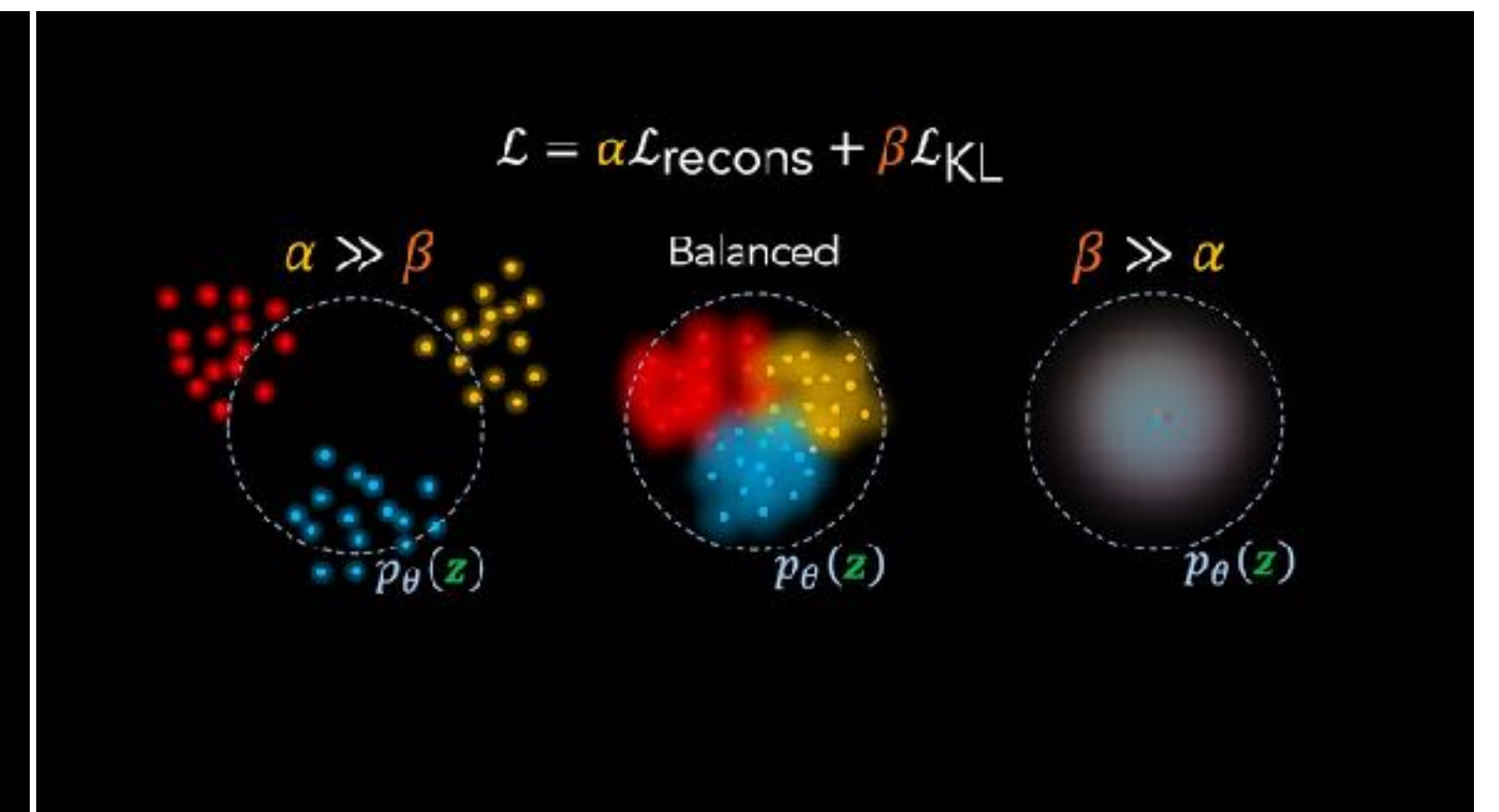
Videóajánló



VAE



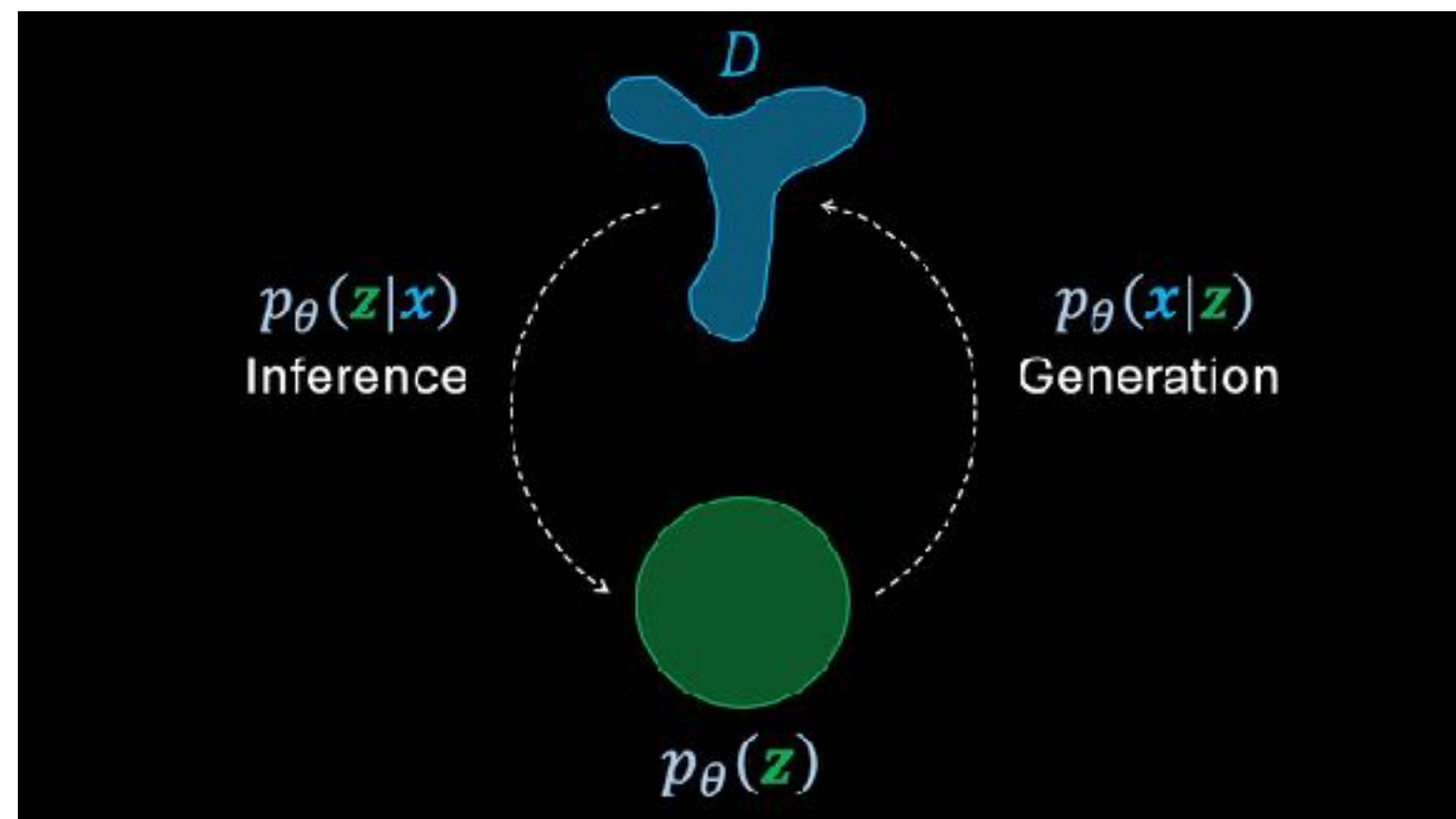
VQVAE



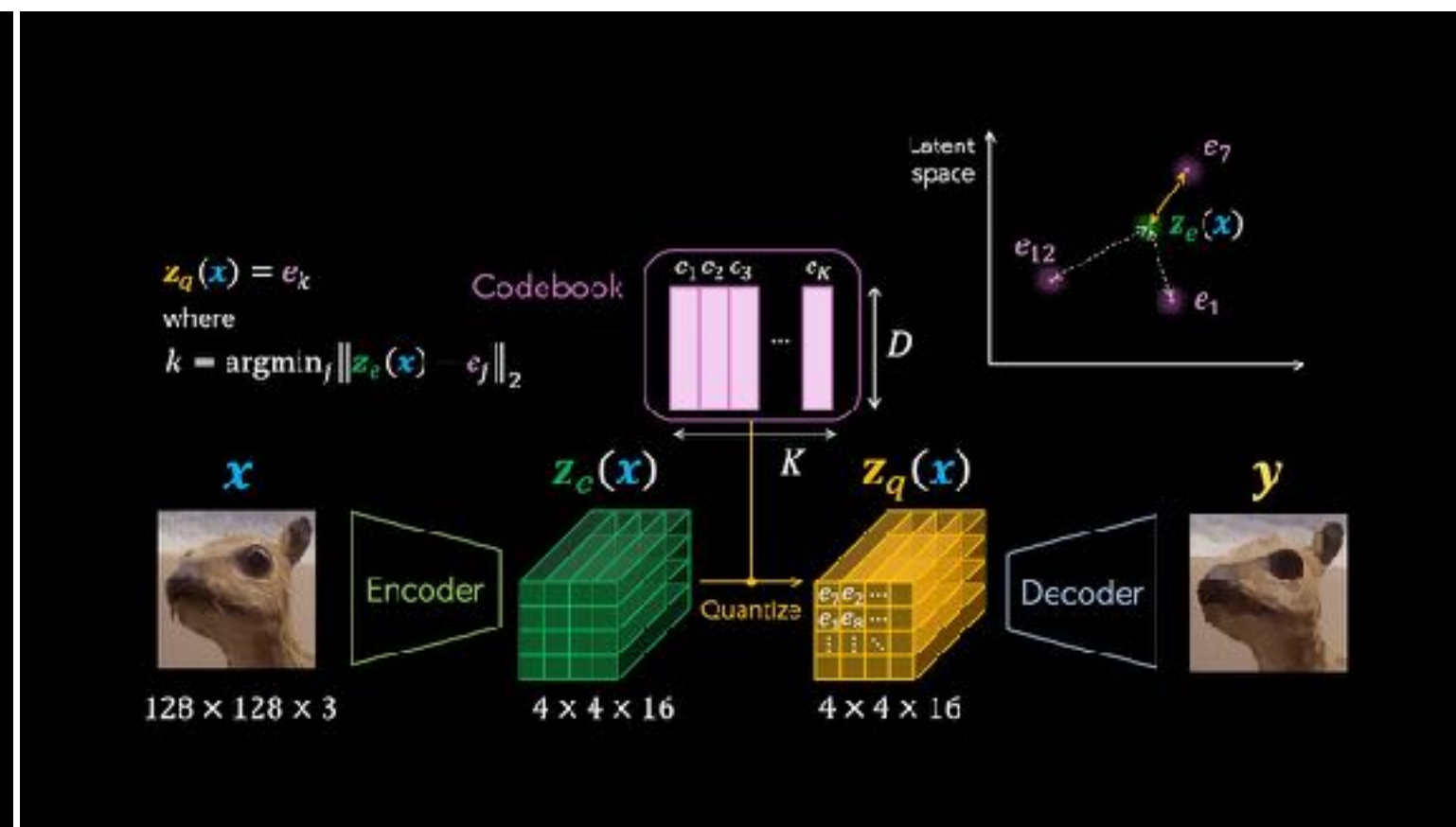
β -VAE

Variációs Autoenkóder (VAE)

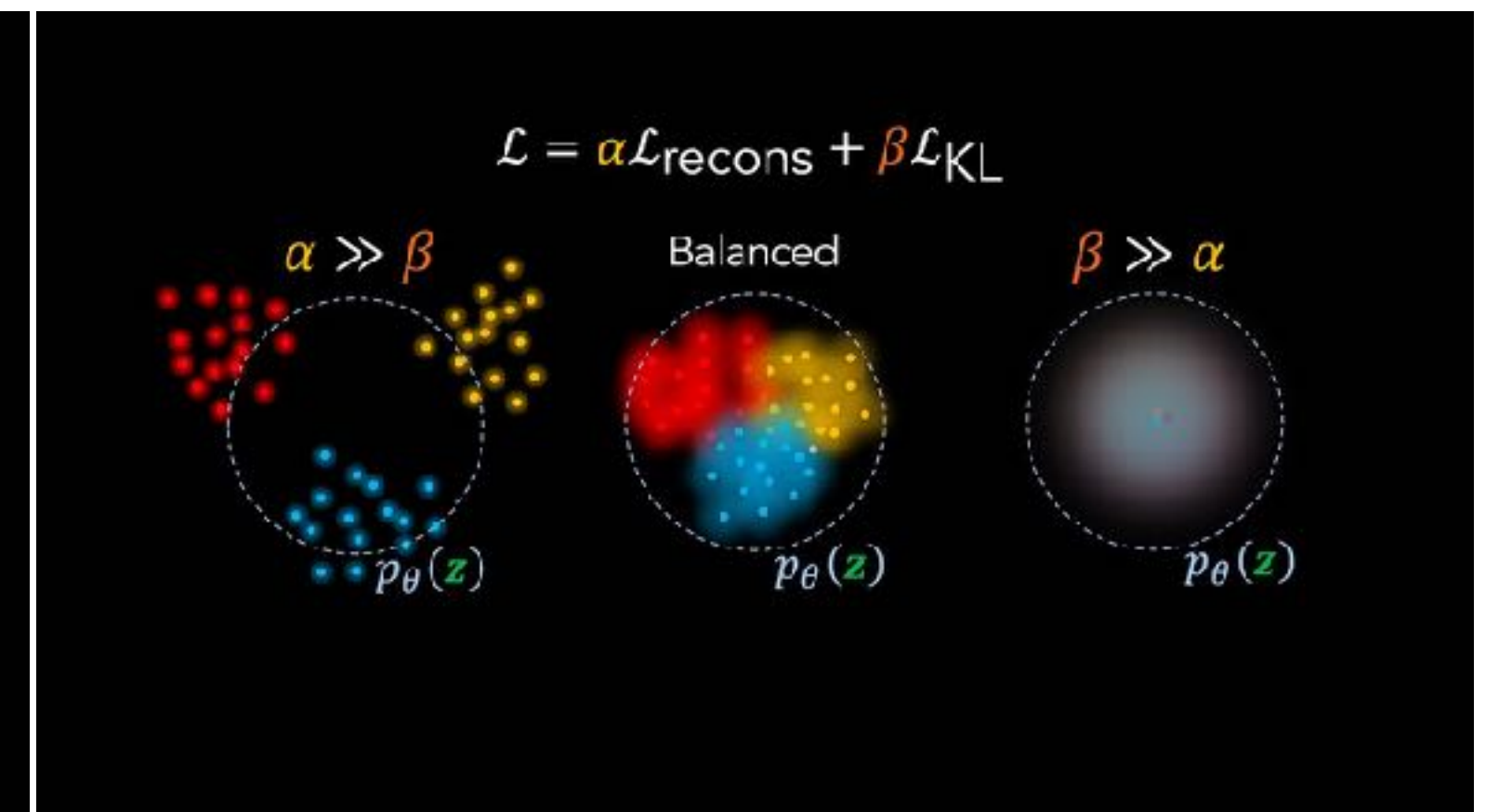
Videóajánló



VAE



VQVAE

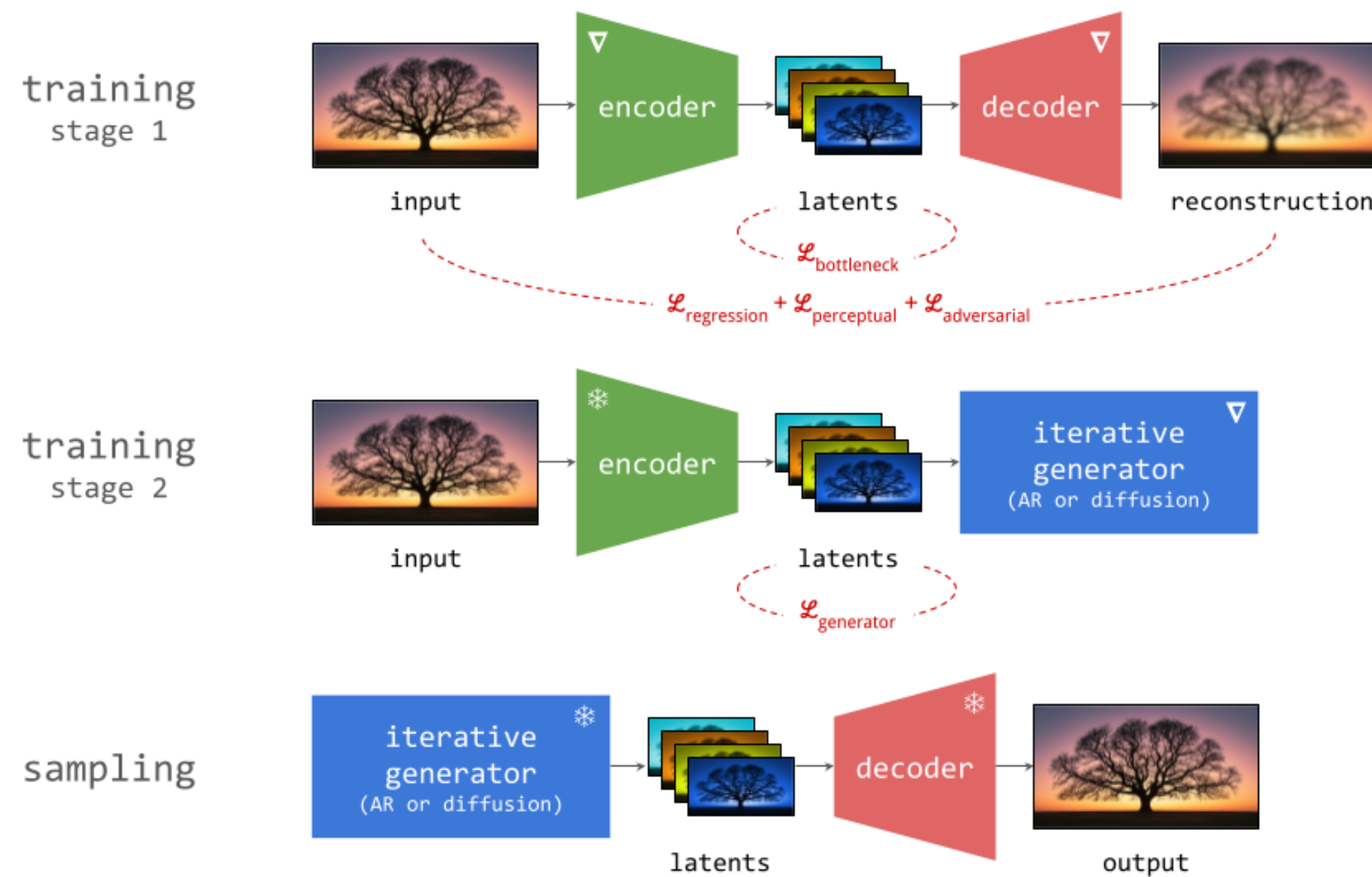


β -VAE

Variációs Autoenkóder (VAE)

Autoenkóderek jelentősége

<https://sander.ai/2025/04/15/latents.html>



Sok generatív modell látens térben operál, külön tanított enkóderrel/dekóderrel!

Variációs Autoenkóder (VAE)

Eredmények



NVAE: A Deep Hierarchical Variational Autoencoder

Arash Vahdat, Jan Kautz
NVIDIA



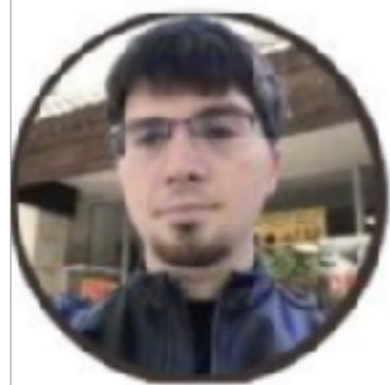
Okosabb (hierarchikus) prior eloszlással

Egy VAE által generált képek gyakran elmosódottak, csak “átlagos értelemben” jók!

Generative Adversarial Networks (GAN)

Generative Adversarial Nets

Ian J. Goodfellow*, Jean Pouget-Abadie†, Mehdi Mirza, Bing Xu, David Warde-Farley,
Sherjil Ozair‡, Aaron Courville, Yoshua Bengio§
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7



Ian Goodfellow

DeepMind
Verified email at deepmind.com - [Homepage](#)
[Deep Learning](#)



TITLE	CITED BY	YEAR
Generative adversarial networks I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, ... Advances in neural information processing systems 27	72471 *	2014

De már 1990-ben(!):

Making the World Differentiable: On Using Self-Supervised Fully Recurrent Neural Networks for Dynamic Reinforcement Learning and Planning in Non-Stationary Environments

Jürgen Schmidhuber*



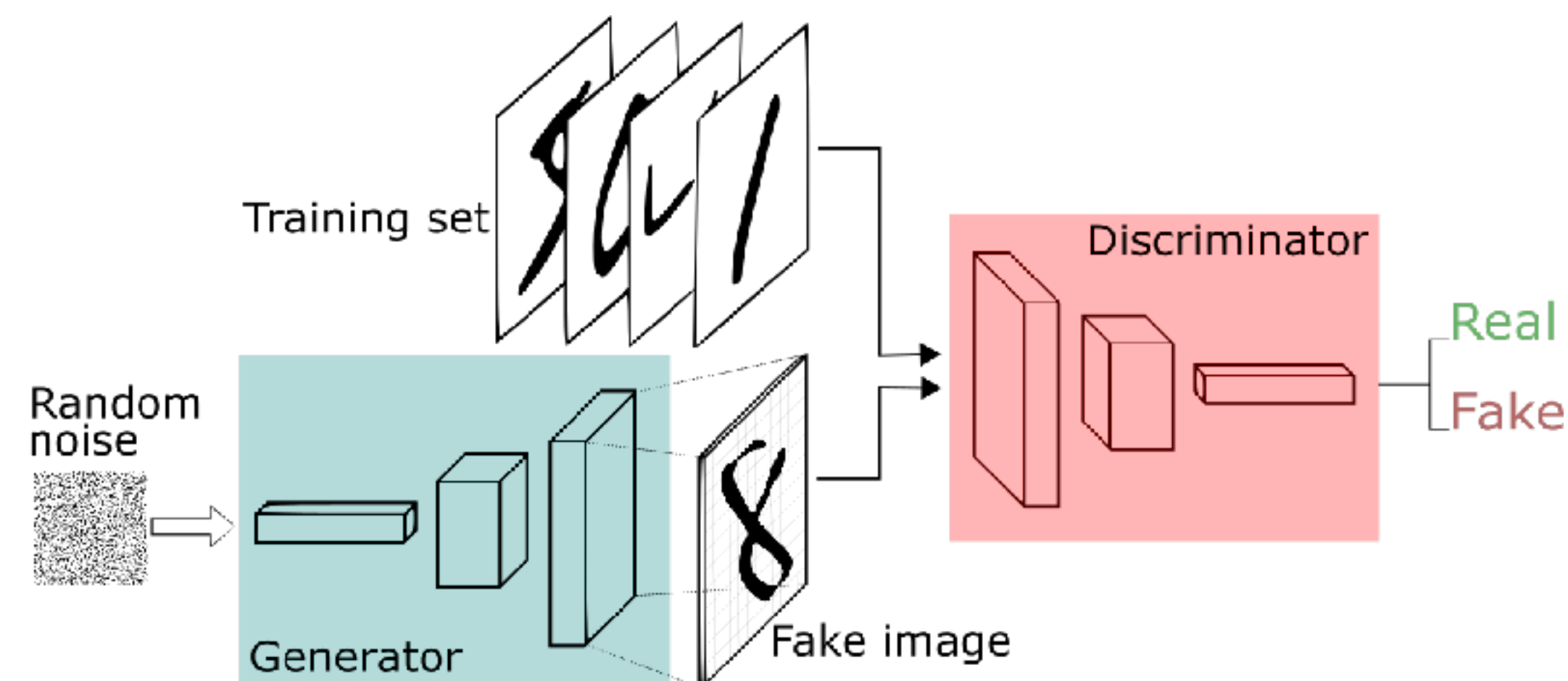
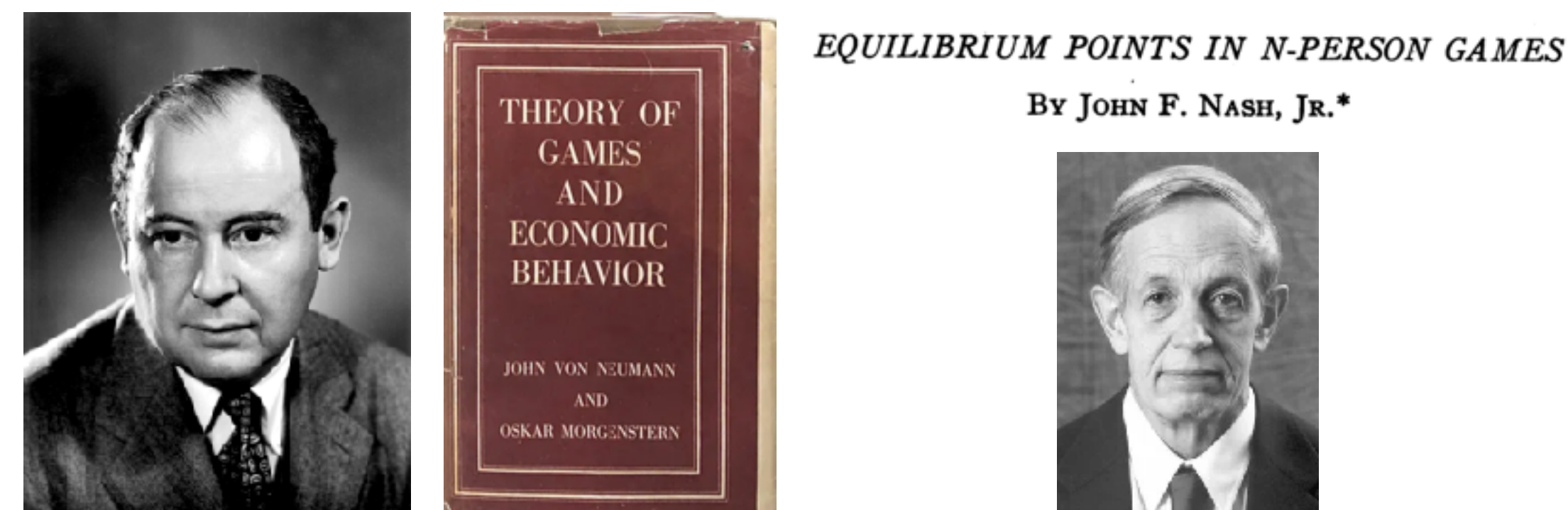
Generative Adversarial Networks (GAN)

GAN loss

- Két háló versenyez egymással (ld. játékelmélet):
 - Generátor $G(z)$: a z látens kódból új x mintát generál.
 - Diszkriminátor $D(x)$: osztályozza az x mintát (Valódi/Hamis).
- Versengő (adversarial) veszteségfüggvény:

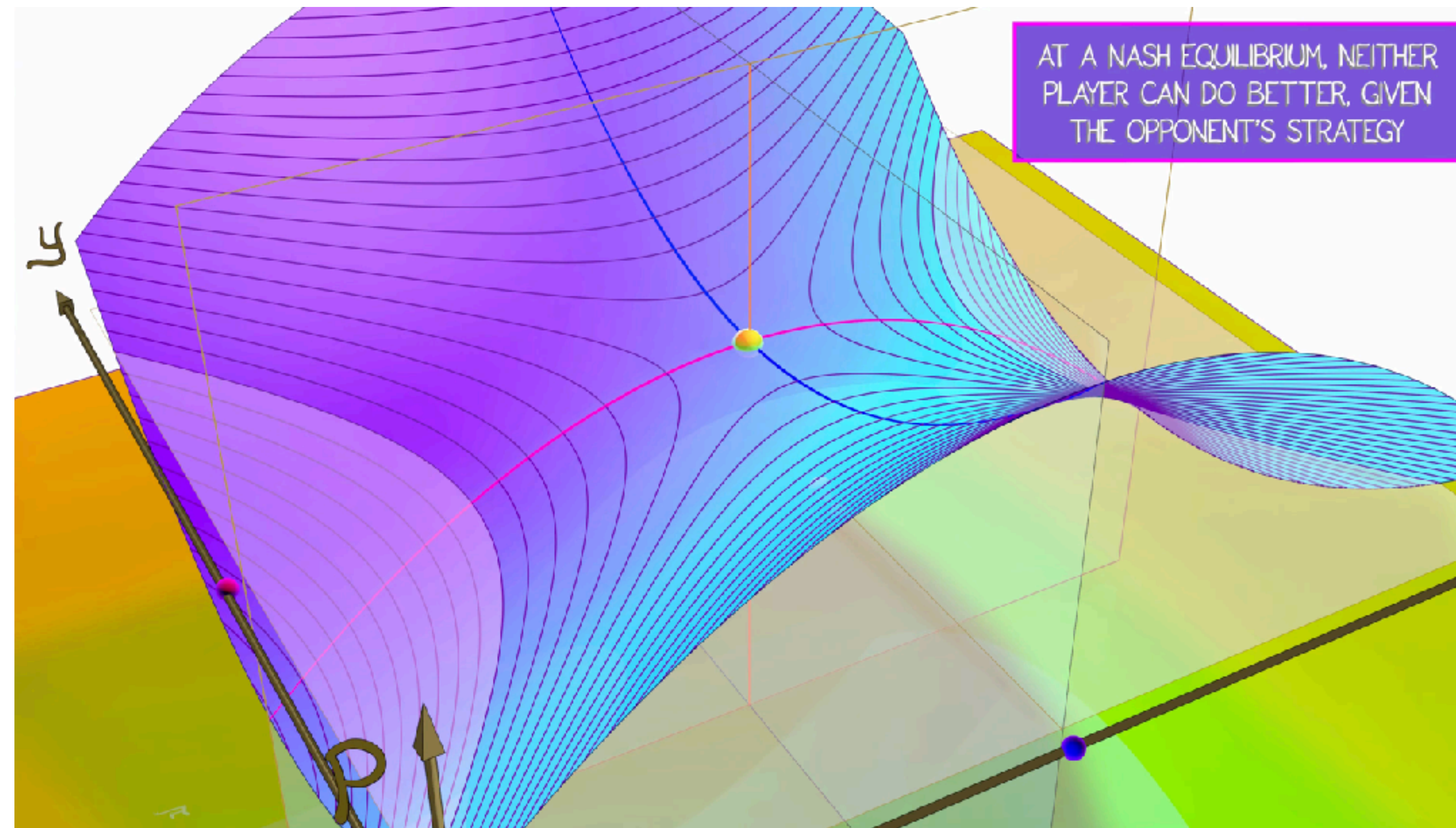
$$\min_D \max_G [\mathbb{E}_x[\log D(x)] + \mathbb{E}_z[1 - \log D(G(z))]]$$

Egyensúlyi (nyereg) pontot keresünk!



Generative Adversarial Networks (GAN)

Nash egyensúly

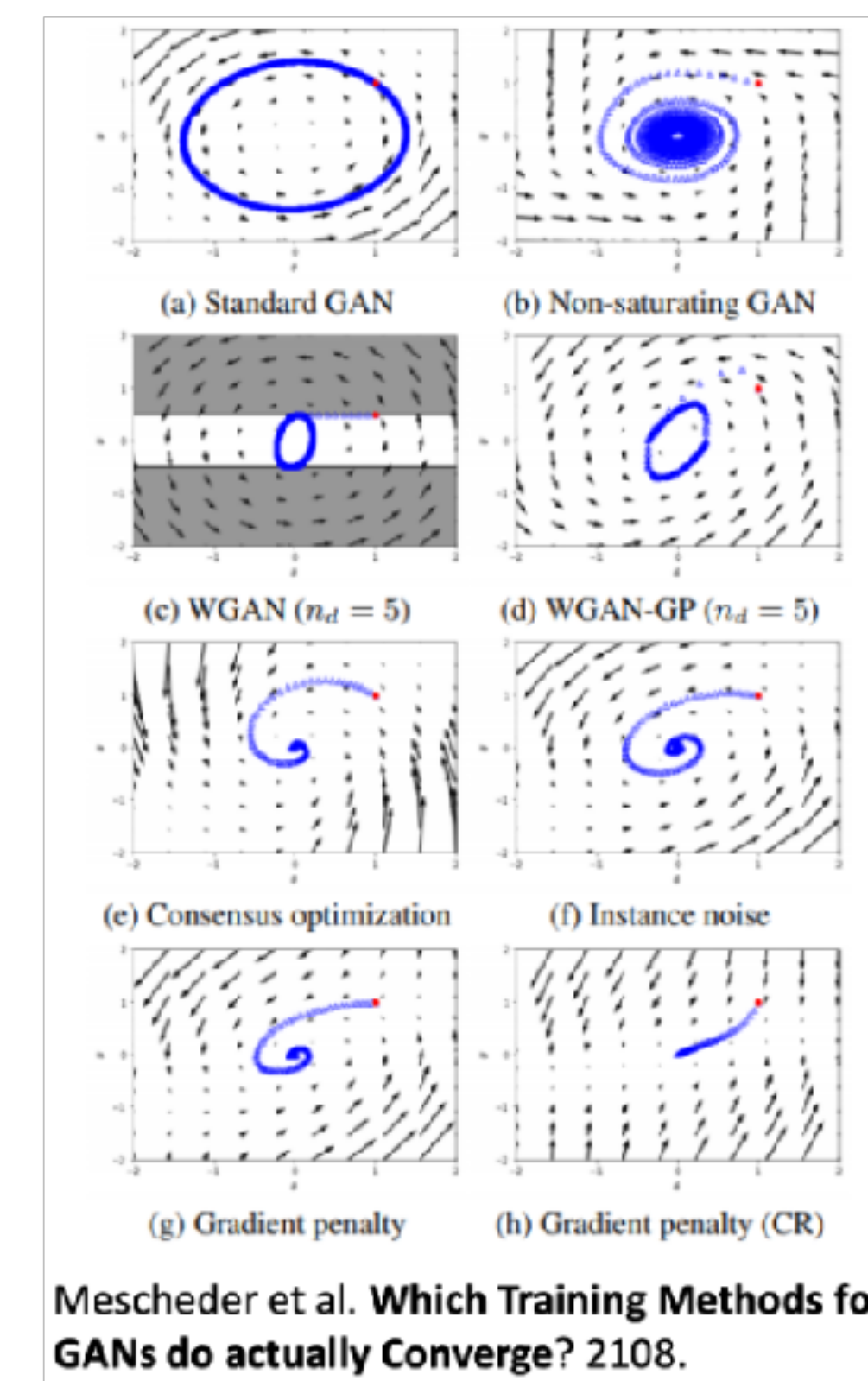
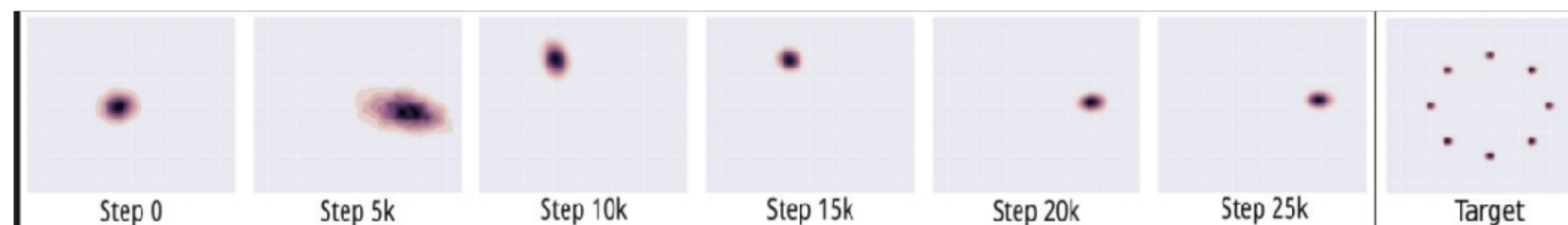


<https://www.youtube.com/watch?v=Moe7EBeUjVc>

Generative Adversarial Networks (GAN)

GAN tanítás

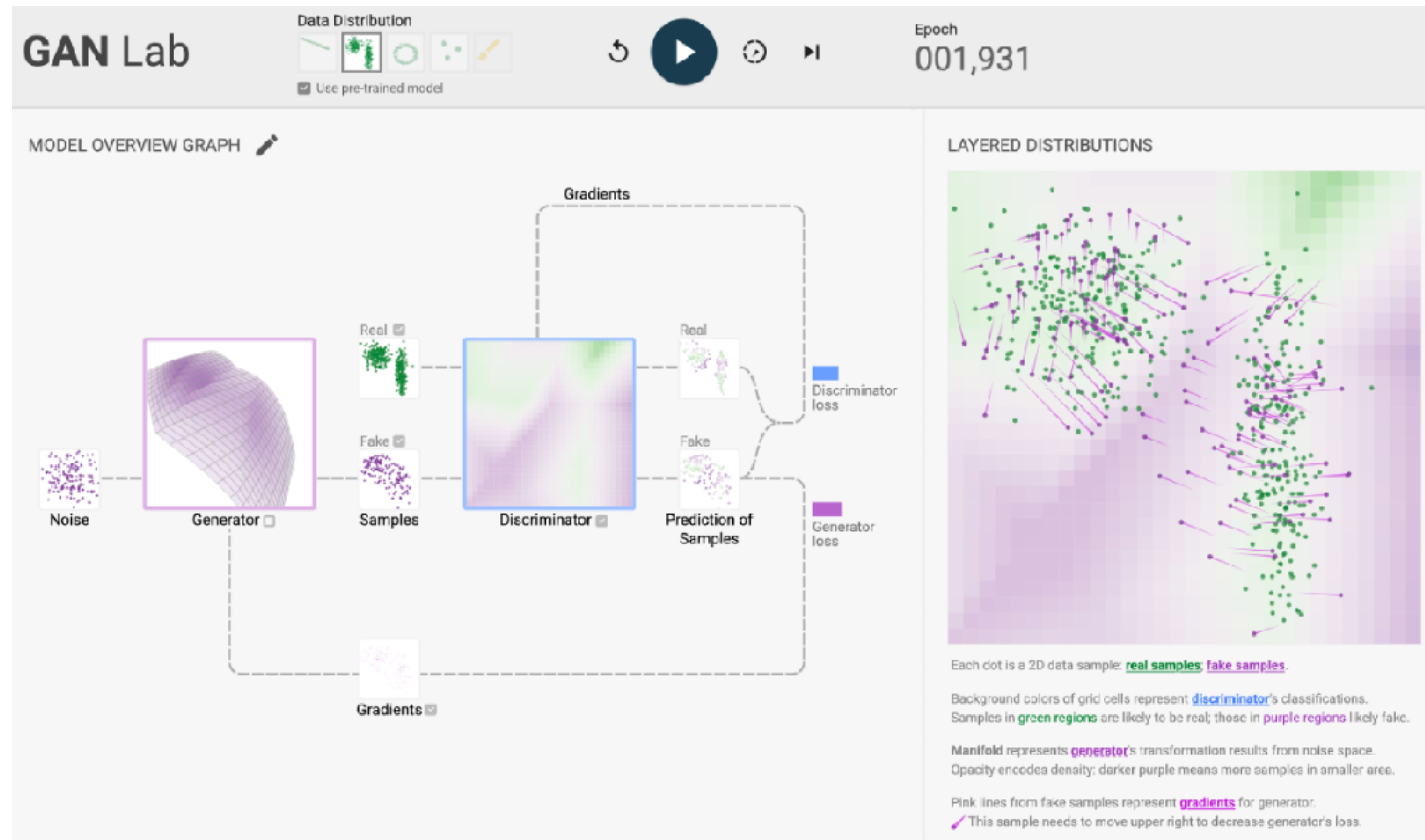
- GAN tanítás — egy ilyen loop vajon sikerre vezet?
 1. Optimalizáljuk a Diszkriminátort
 2. Optimalizáljuk a Generátort
- A GAN hálók tanítása sajnos *legendásan nehéz*...
 - Egyik háló sem nyerheti meg a játékot!
 - Ha a Diszkriminátor 100%-ra teljesít a tanító halmazon, a Generátor az egyik módusra tanul rá (**mode collapse**)



Rengeteg különböző módszer, loss variáció, stb...

Generative Adversarial Networks (GAN)

GAN Lab



<https://poloclub.github.io/ganlab/>

Generative Adversarial Networks (GAN)

Fejlődés



Ian Goodfellow
@goodfellow_ian

4.5 years of GAN progress on face generation.

arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434

arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196

arxiv.org/abs/1812.04948



Az első igazán “ütős” generatív modellek GAN-ok voltak!

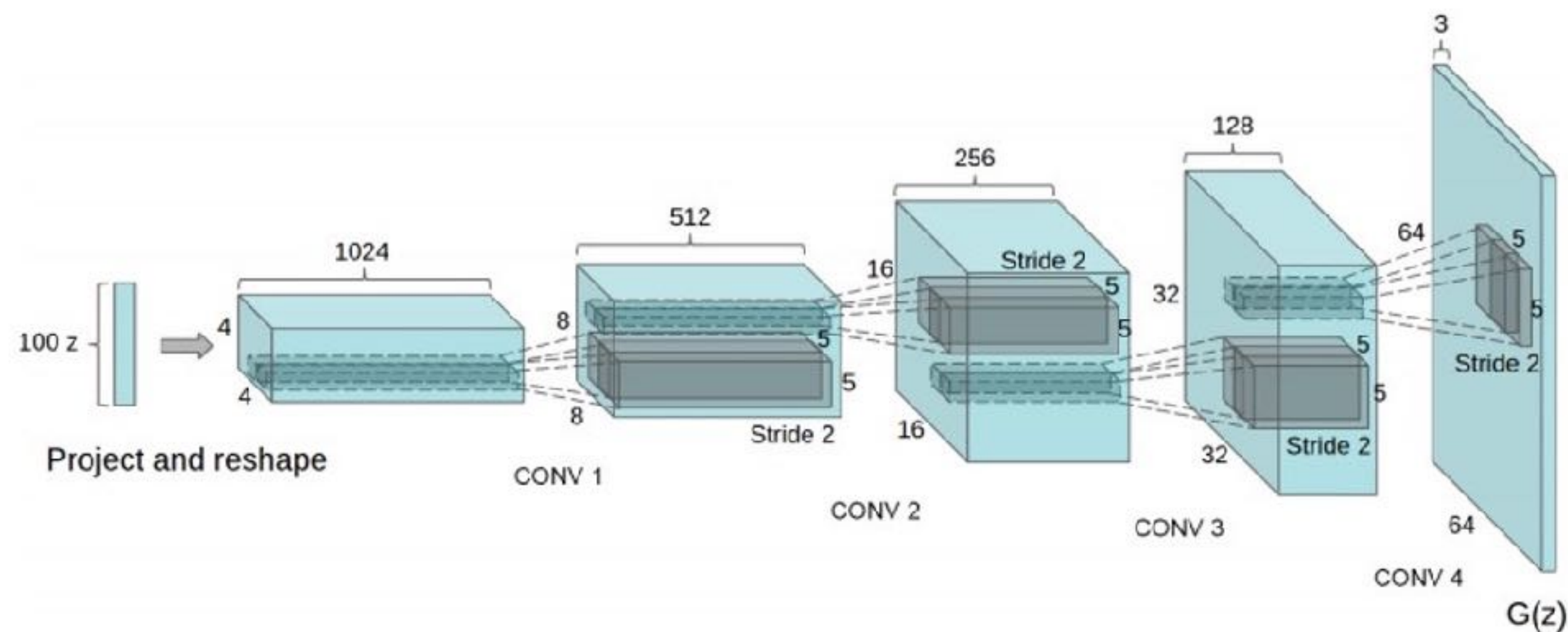
Generative Adversarial Networks (GAN)

Deep Convolutional GAN (DCGAN) [2016]

UNSUPERVISED REPRESENTATION LEARNING
WITH DEEP CONVOLUTIONAL
GENERATIVE ADVERSARIAL NETWORKS

Alec Radford & Luke Metz
indico Research
Boston, MA
{alec, luke}@indico.io

Soumith Chintala
Facebook AI Research
New York, NY
soumith@fb.com



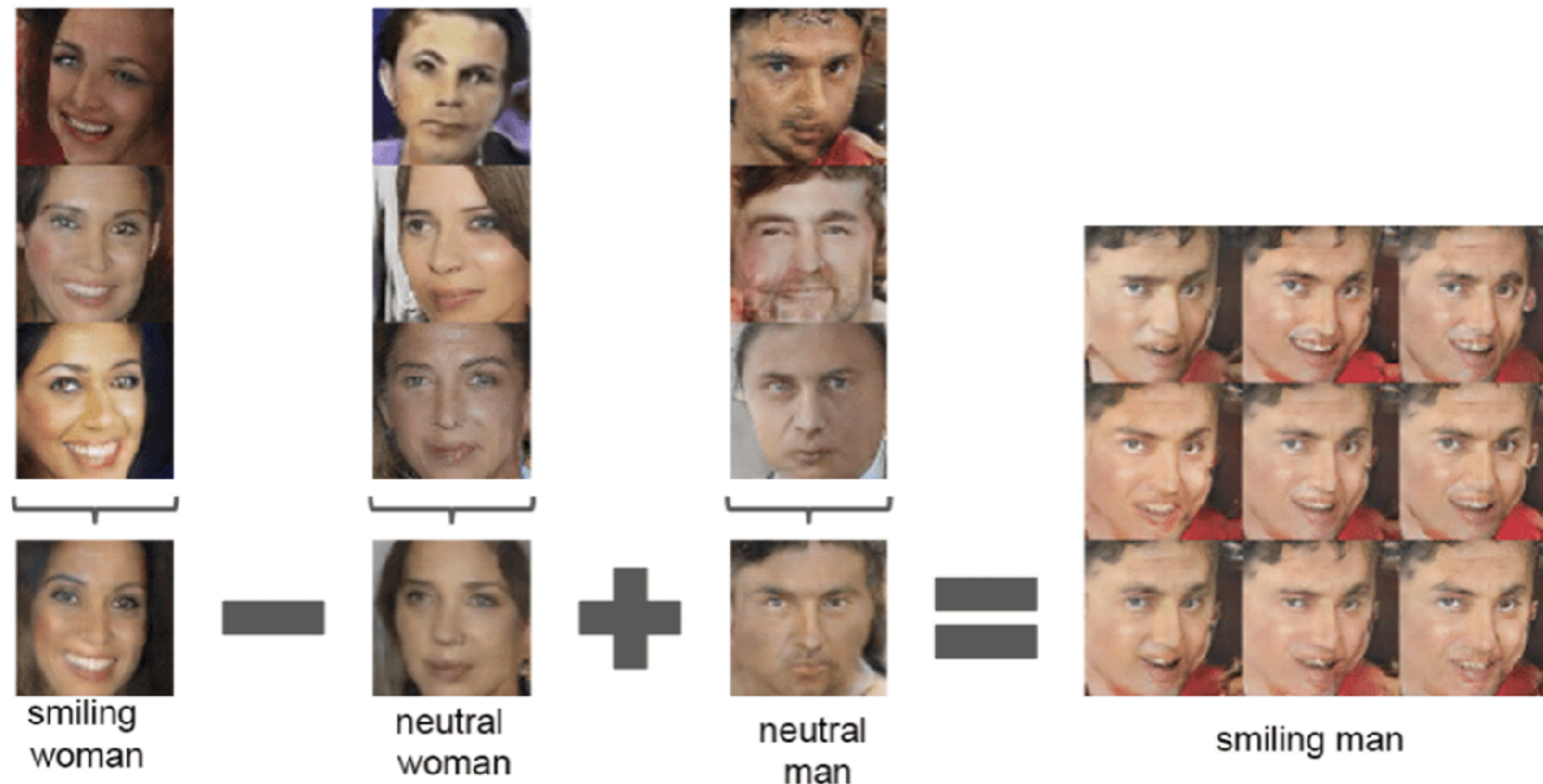
Generative Adversarial Networks (GAN)

DCGAN – Látens tér interpoláció



Generative Adversarial Networks (GAN)

DCGAN – Látens tér aritmetika



Generative Adversarial Networks (GAN)

Progresszív GAN [2017]

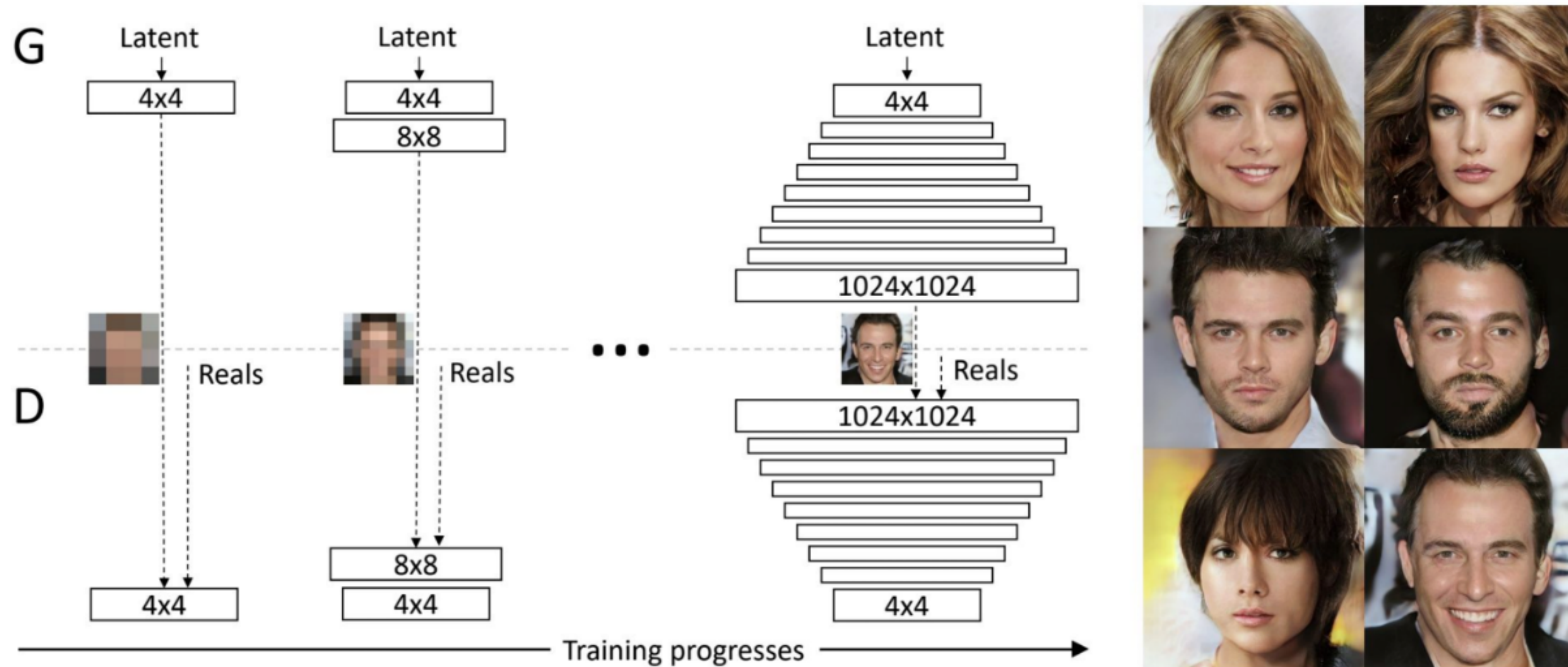
PROGRESSIVE GROWING OF GANS FOR IMPROVED QUALITY, STABILITY, AND VARIATION

Tero Karras
NVIDIA

Timo Aila
NVIDIA

Samuli Laine
NVIDIA

Jaakko Lehtinen
NVIDIA and Aalto University



Ötlet: tanítsunk fokozatosan, növekvő felbontású képeken!

Generative Adversarial Networks (GAN)

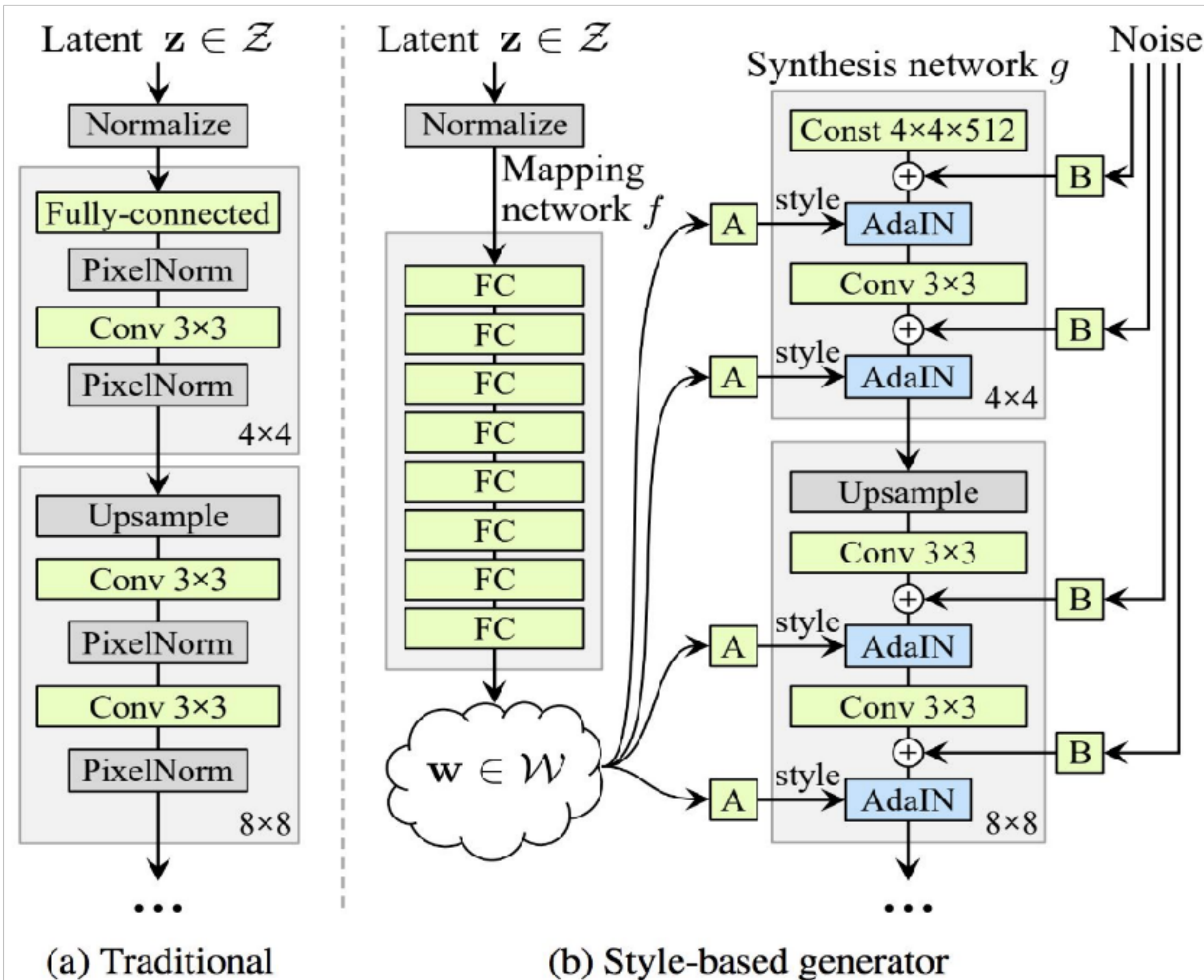
StyleGAN [2018]

A Style-Based Generator Architecture for Generative Adversarial Networks

Tero Karras
NVIDIA

Samuli Laine
NVIDIA

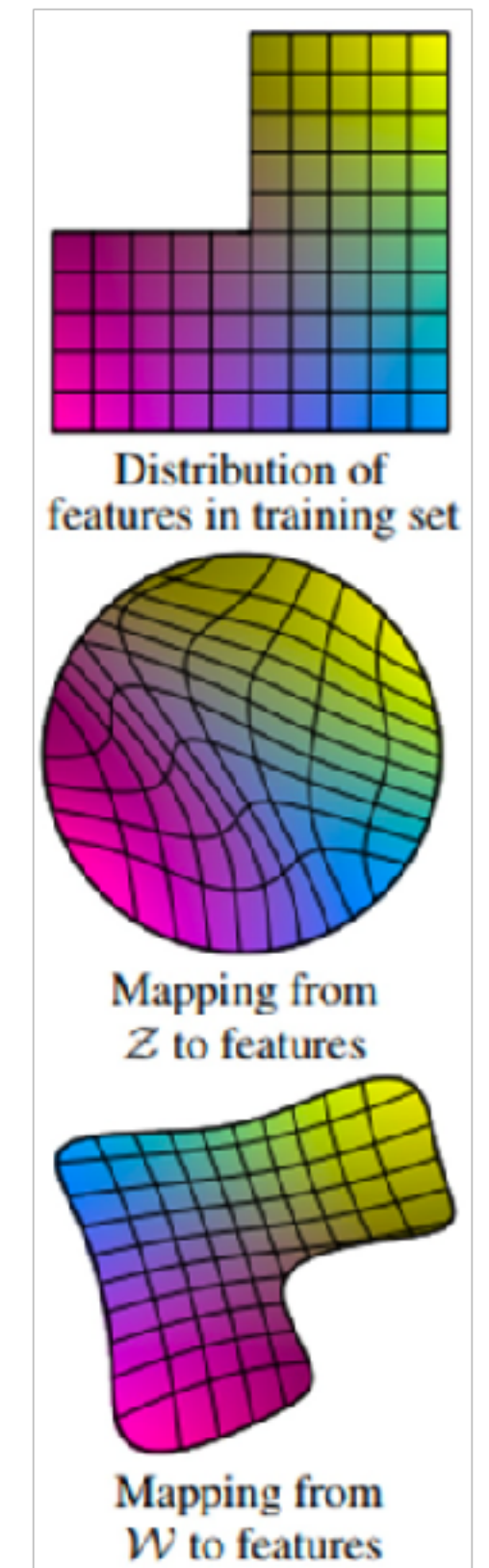
Timo Aila
NVIDIA



1. Minden felbontást modulál (AdaIN) egy stílusvektor alapján
2. Minden felbontáson zajt visz be (a la diffúzió!)

Adaptív Instance Norm:

$$\text{AdaIN}(x_i, y) = y_{s,i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i}$$



Generative Adversarial Networks (GAN)

StyleGAN – Eredmények



Generálás nagy felbontáson (1024 x 1024)!

<https://thispersondoesnotexist.com/>

<https://thisanimedoesnotexist.ai/>

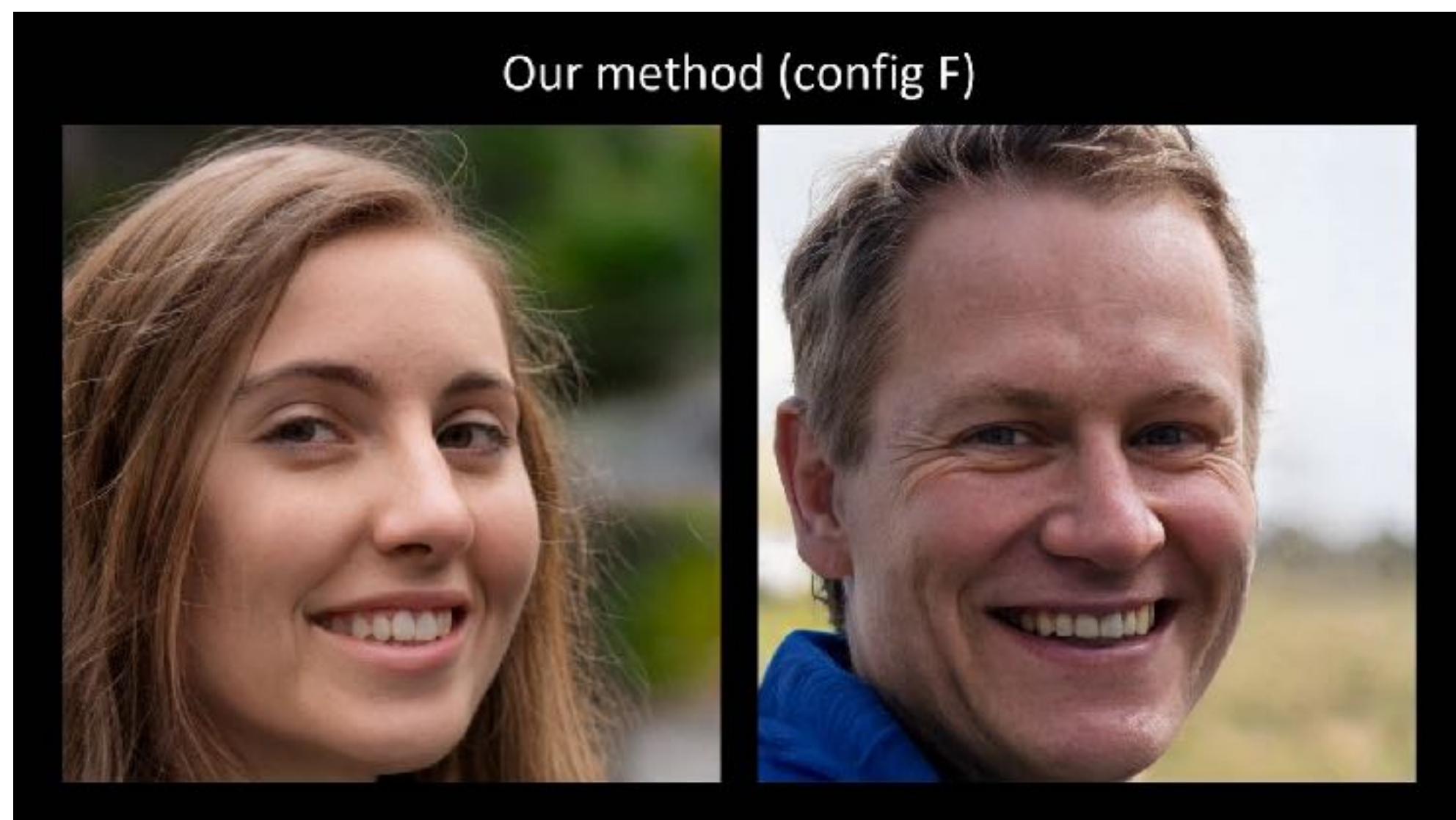
Generative Adversarial Networks (GAN)

StyleGAN – Stílus átvitel



Generative Adversarial Networks (GAN)

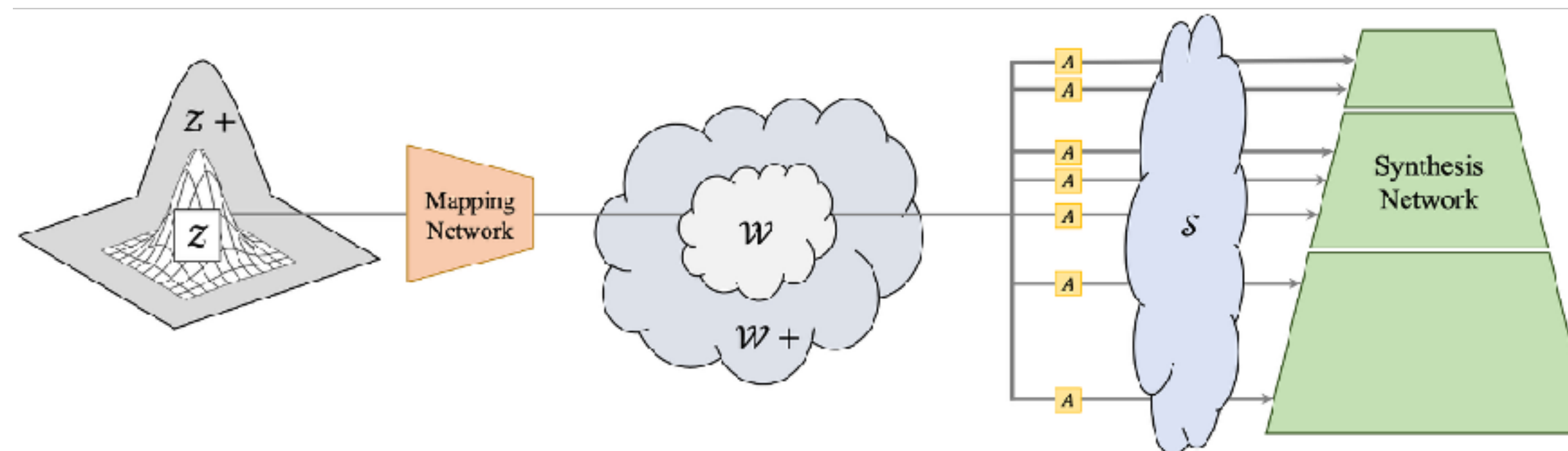
StyleGAN – Variánsok*



StyleGAN2

State-of-the-Art in the Architecture, Methods and Applications of StyleGAN

AMIT H. BERMANO, Tel Aviv University
RINON GAL, Tel Aviv University
YUVAL ALALUF, Tel Aviv University
RON MOKADY, Tel Aviv University
YOTAM NITZAN, Tel Aviv University
OMER TOV, Tel Aviv University
OR PATASHNIK, Tel Aviv University
DANIEL COHEN-OR, Tel Aviv University



Generative Adversarial Networks (GAN)

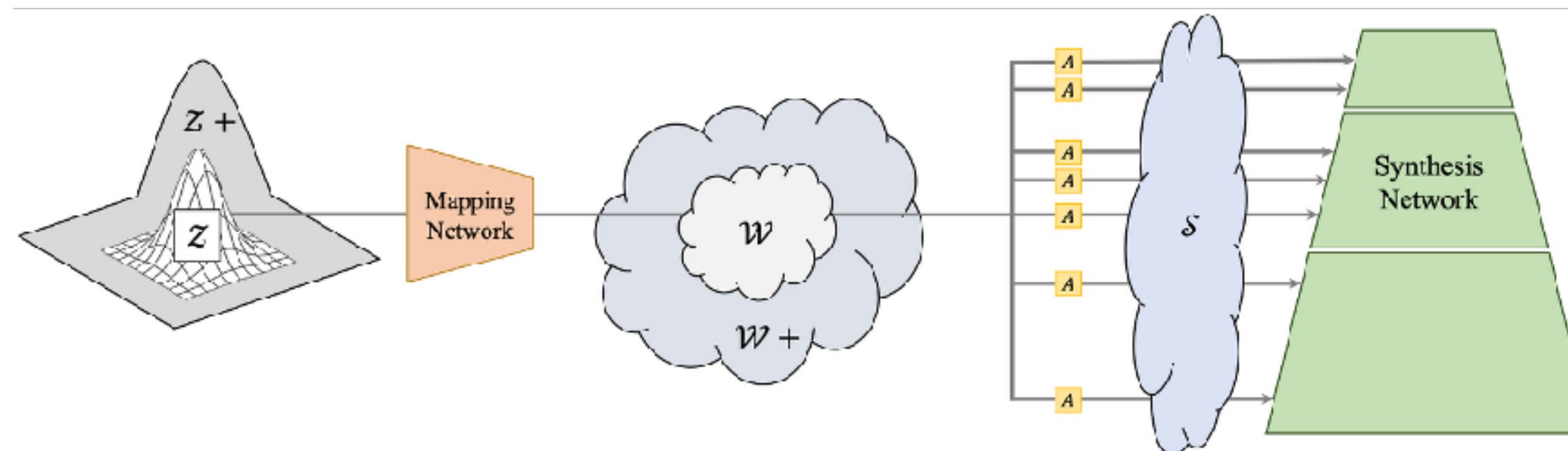
StyleGAN – Variánsok*



StyleGAN2

State-of-the-Art in the Architecture, Methods and Applications of StyleGAN

AMIT H. BERMANO, Tel Aviv University
RINON GAL, Tel Aviv University
YUVAL ALALUF, Tel Aviv University
RON MOKADY, Tel Aviv University
YOTAM NITZAN, Tel Aviv University
OMER TOV, Tel Aviv University
OR PATASHNIK, Tel Aviv University
DANIEL COHEN-OR, Tel Aviv University



Generative Adversarial Networks (GAN)

Sic transit gloria mundi...

- 2020-tól beindult a diffúziós modellek fejlődése...
- A GAN alapú módszereket hanyagolni kezdték...
- Vannak ellenvélemények:

GANs DIDN'T FAIL, THEY WERE ABANDONED

GAN, AI scaling

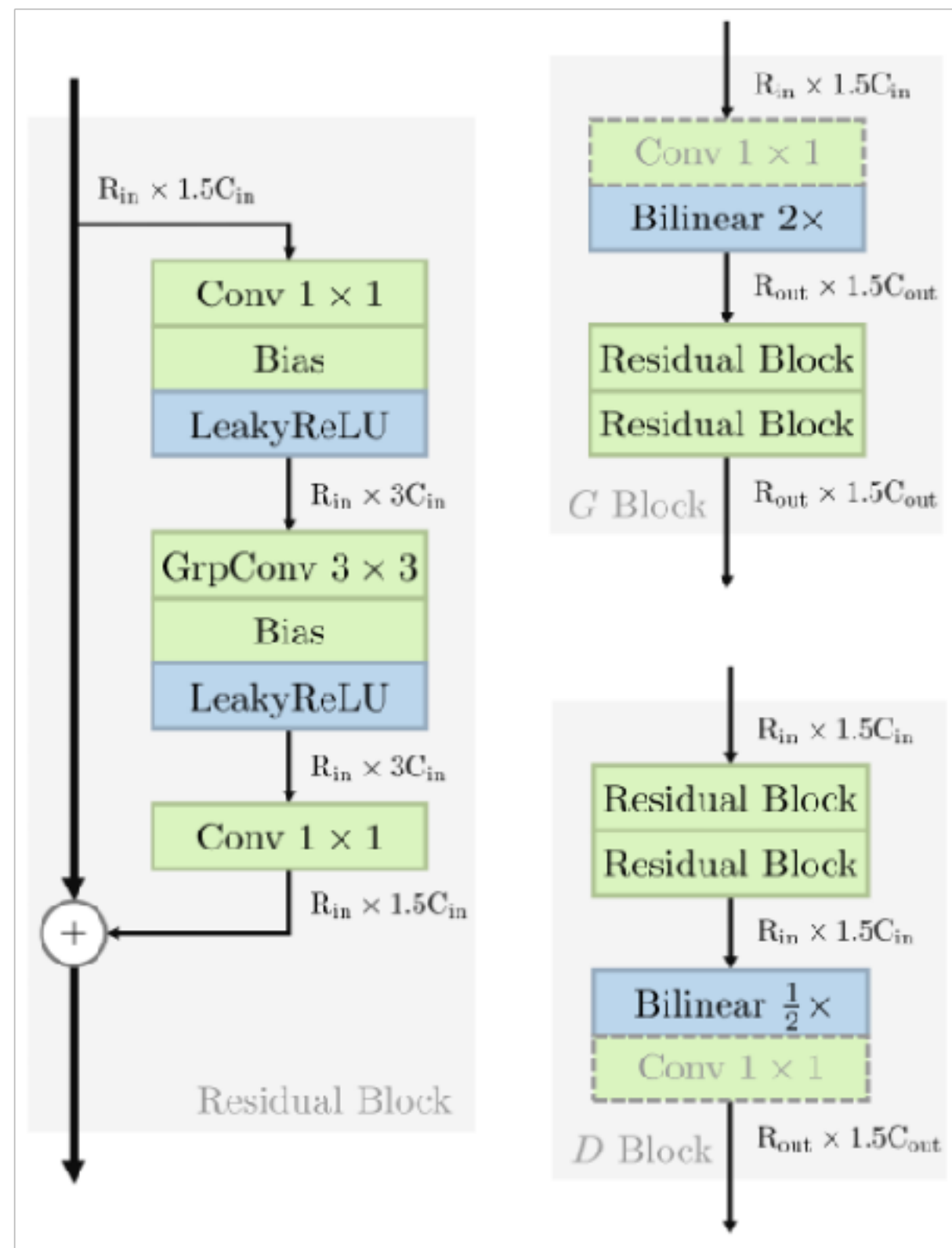
Diffusion models supposedly beat GANs because they scale better and stabler. That is unproven, and false. GANs should be revisited.

by: *Gwern Branwen* · 2022-10-04–2022-10-15 · *notes* ○ · *certainty: possible* ○ · *importance: 4*

<https://gwern.net/gan>

Generative Adversarial Networks (GAN)

R3GAN*



GAN Will Return In Avengers: Doomsday ?

- Modernabb architektúrális elemek (ResNet)
- Új loss
- Transformer hálóktól kölcsönzött trükkök

The GAN is dead; long live the GAN! A Modern Baseline GAN

Yiwen Huang
Brown University

Aaron Gokaslan
Cornell University

Volodymyr Kuleshov
Cornell University

James Tompkin
Brown University

Configuration	FID↓	G #params	D #params
A StyleGAN2	7.516	24.767M	24.001M
B Stripped StyleGAN2			
- z normalization			
- Minibatch stddev			
- Equalized learning rate			
- Mapping network			
- Style injection			
- Weight mod / demod	12.46	18.890M	23.996M
- Noise injection			
- Mixing regularization			
- Path length regularization			
- Lazy regularization			
C Well-behaved Loss			
+ RpGAN loss	11.77	18.890M	23.996M
+ R_2 gradient penalty	11.65	18.890M	23.996M
D ConvNeXt-ify pt. 1			
+ ResNet-ify G & D	10.17	23.400M	23.282M
- Output skips	9.950	23.378M	23.282M
E ConvNeXt-ify pt. 2			
+ ResNeXt-ify G & D	7.507	23.188M	23.091M
+ Inverted bottleneck	7.045	23.058M	23.010M


Table 2: Effect of our simplification and modernization efforts evaluated on FFHQ-256.

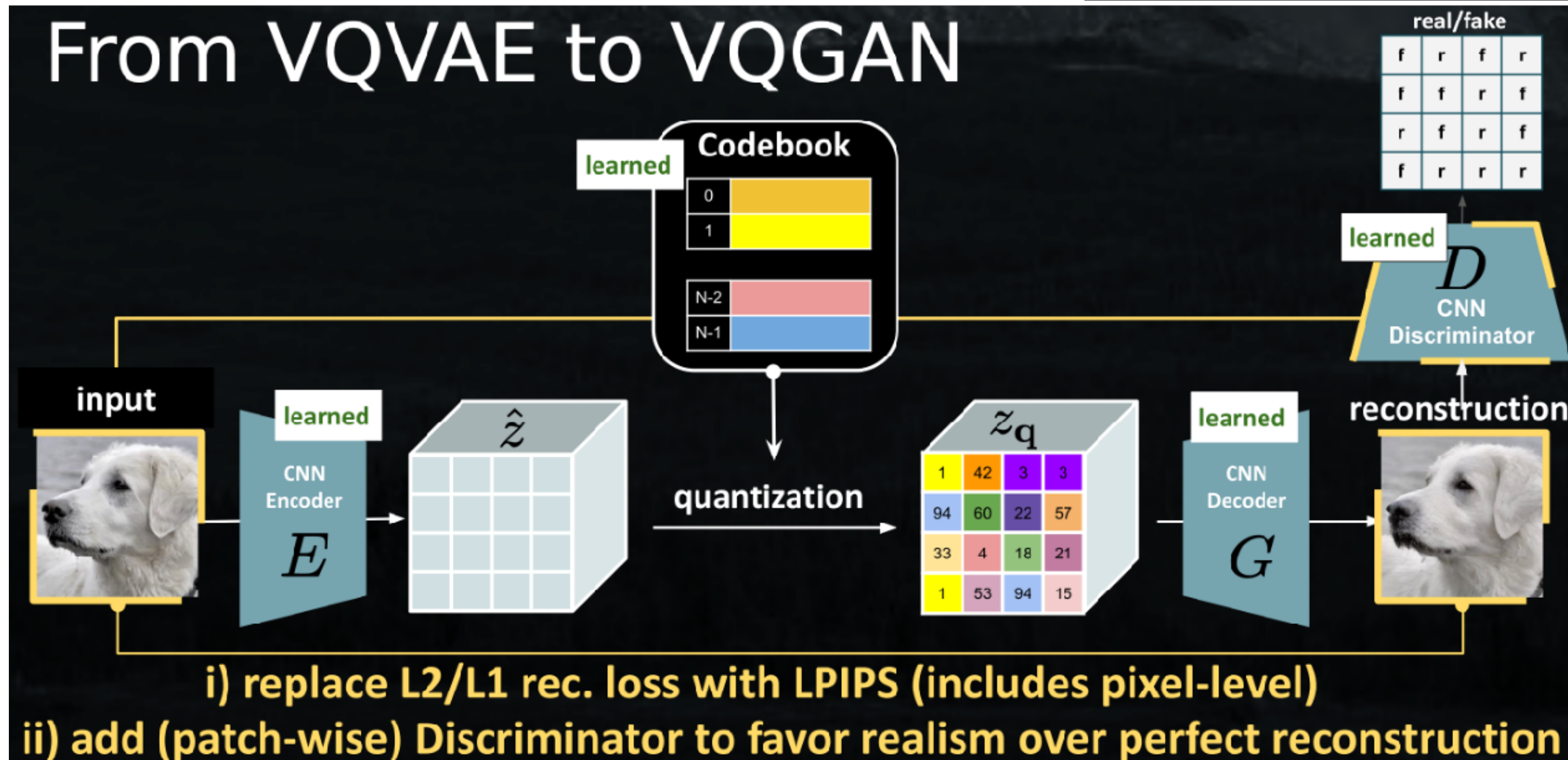
Generative Adversarial Networks (GAN)

VQGAN

Taming Transformers for High-Resolution Image Synthesis

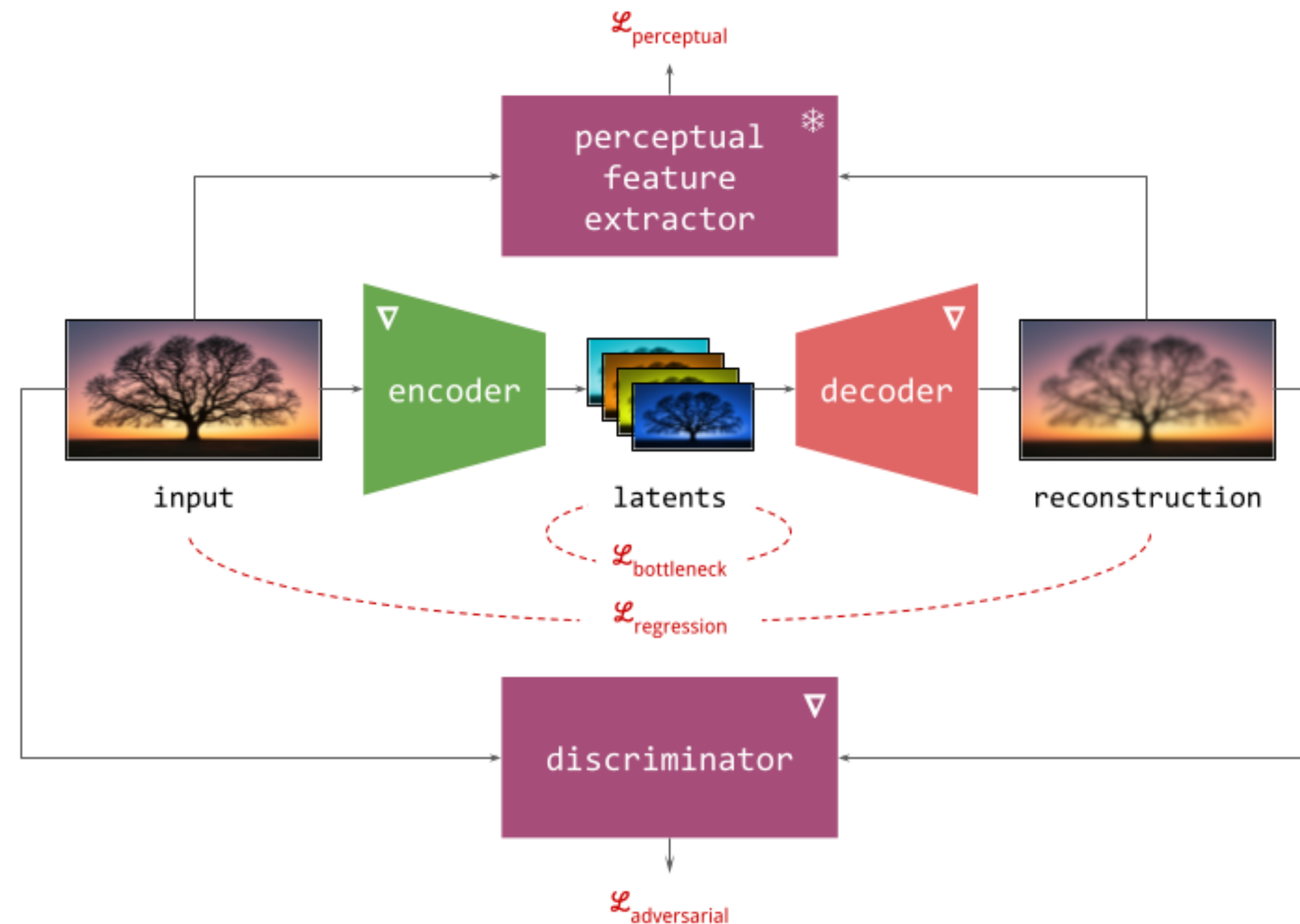
Patrick Esser* Robin Rombach* Björn Ommer
Heidelberg Collaboratory for Image Processing, IWR, Heidelberg University, Germany

Sander Dieleman  @sedielem
IMO VQGAN is why GANs deserve the NeurIPS test of time award. Suddenly our image representations were an order of magnitude more compact. Absolute game changer for generative modelling at scale, and the basis for latent diffusion models.



Generative Adversarial Networks (GAN)

GAN but not forgotten



A látens dekódereket / tokenizálókat manapság is GAN loss-szal tanítják!

Következő előadás: Diffúziós Generatív Modellek

- Zajcsökkentő autoenkóderek
- Diffúziós modellek alapjai
- Folyamillesztés

