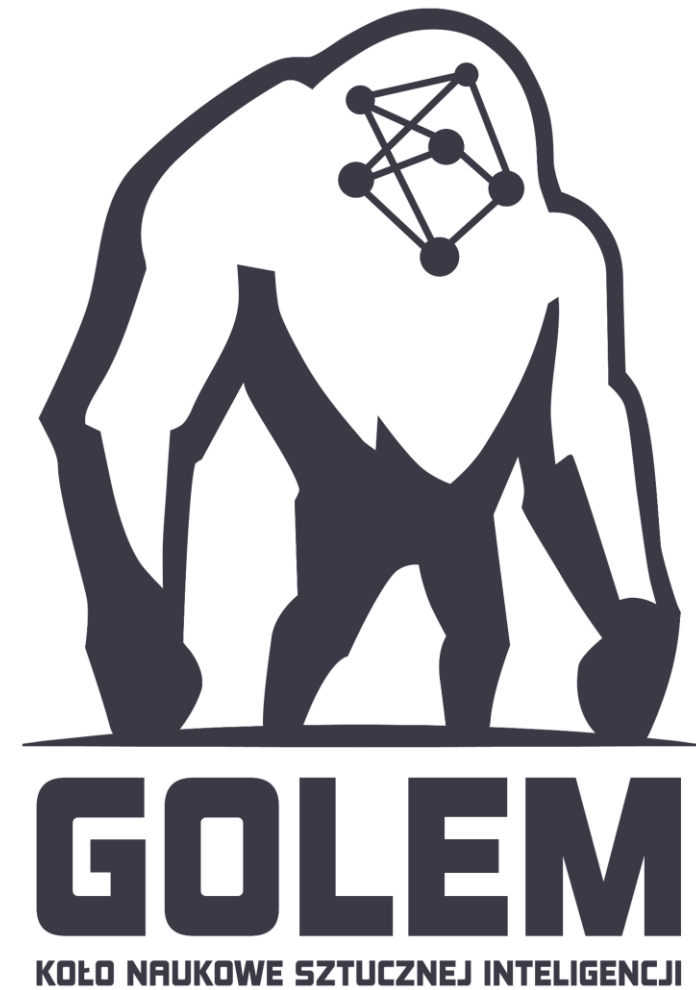


Trochę o sieciach generatywnych

Od VAE do GANa przez
VAEGANa



O czym opowiem?

1. Na czym polegają modele generatywne?
2. Zastosowania modeli generatywnych.
3. Czemu autokoder jako model generatywny nie działa?
4. Dlaczego VAE jako model generatywny działa lepiej.
5. Intuicja stojąca za GANem.
6. VAEGAN = VAE+
7. Kodzik

Modele generatywne

Cel: Nauczyć się rozkładu prawdopodobieństwa, który jak najbardziej przypomina rozkład prawdopodobieństwa danych treningowych.
Następnie próbkując ten rozkład będziemy w stanie generować nowe obserwacje.



Dane treningowe $\sim p_{data}(x)$



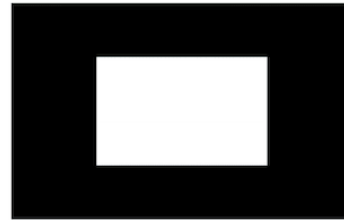
Generacje $\sim p_{model}(x)$

Chcemy, aby $p_{model}(x)$ był jak najbardziej podobny do $p_{data}(x)$.

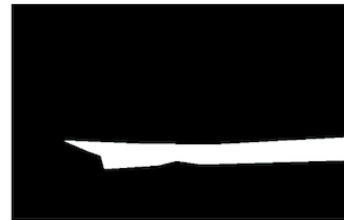
Zastosowania modeli generatywnych



(a)



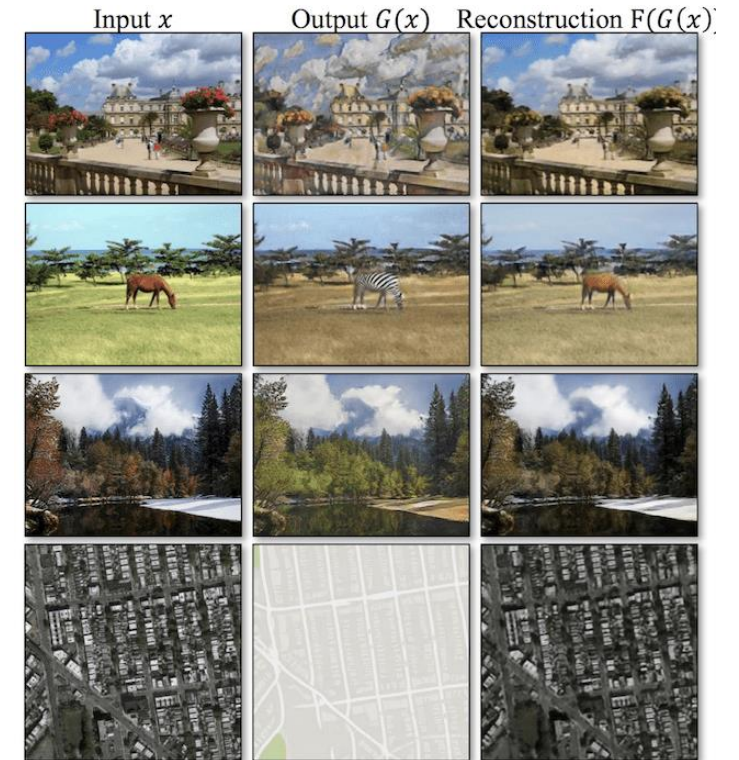
(c)



(b)



(d)



<https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/>

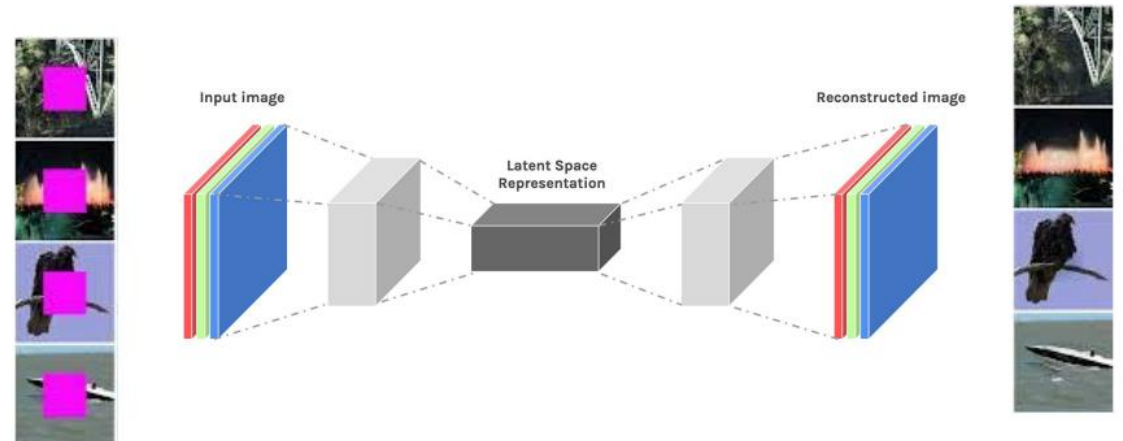
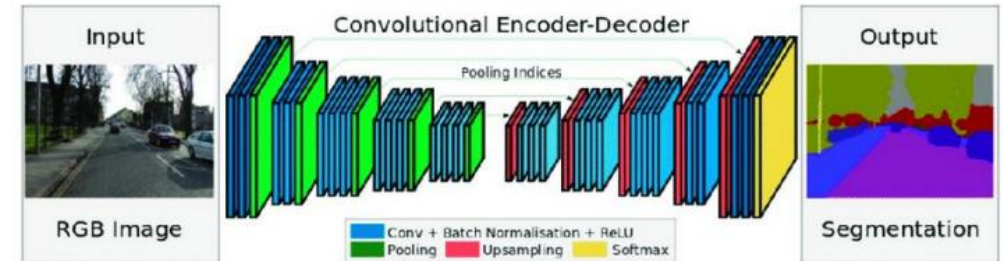
<https://thisxdoesnotexist.com/>

Autokoder

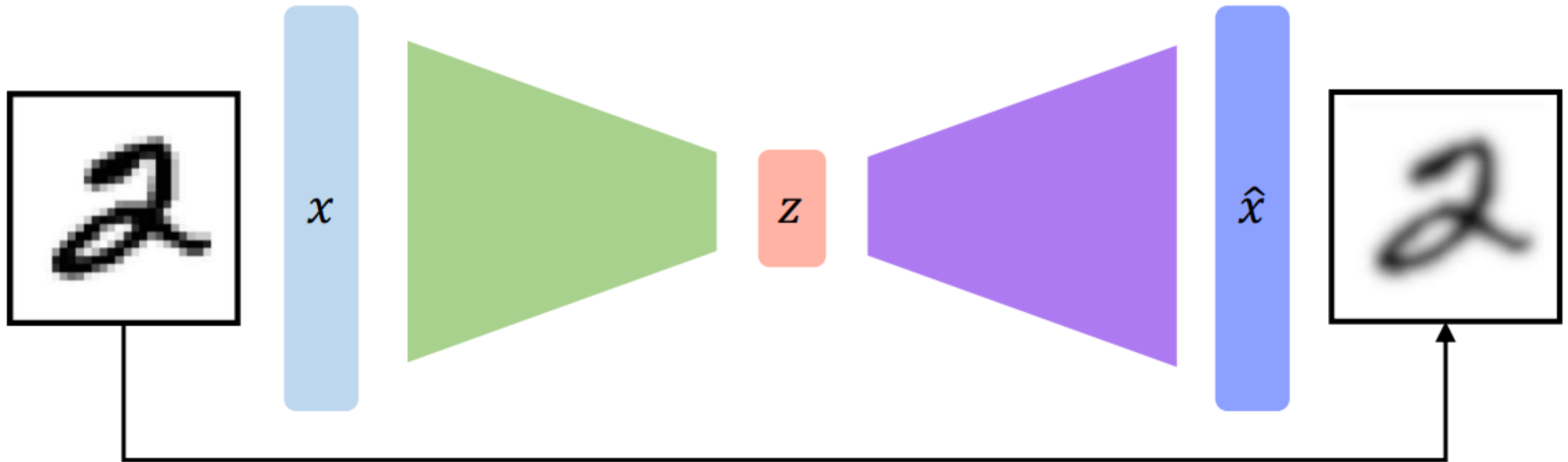
Cel: Nauczyć się funkcji, która ze skompresowanej reprezentacji zrekonstruuje oryginalny obraz.

Zastosowania:

- Kompresja danych
- Segmentacja semantyczna
- „Neural inpainting”
- Usuwanie szumu



Autokoder



http://introtodeeplearning.com/2019/materials/2019_6S191_L4.pdf

Autokoder – funkcja straty

$$L_{AE}(\theta, \phi) = \frac{1}{n} \sum_{i=1}^n (x^i - p_{\theta}(q_{\phi}(x^i)))^2$$

p – dekodek

q – kodek

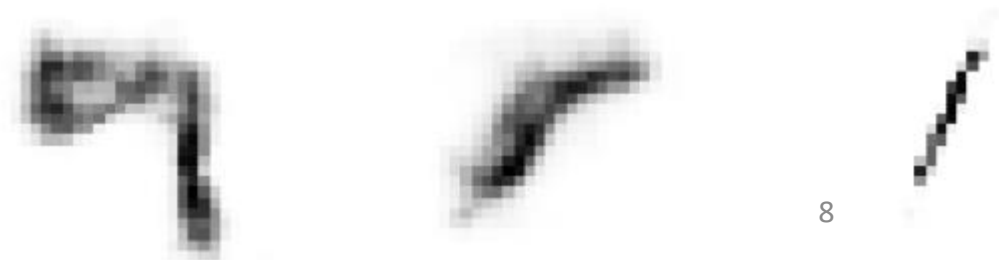
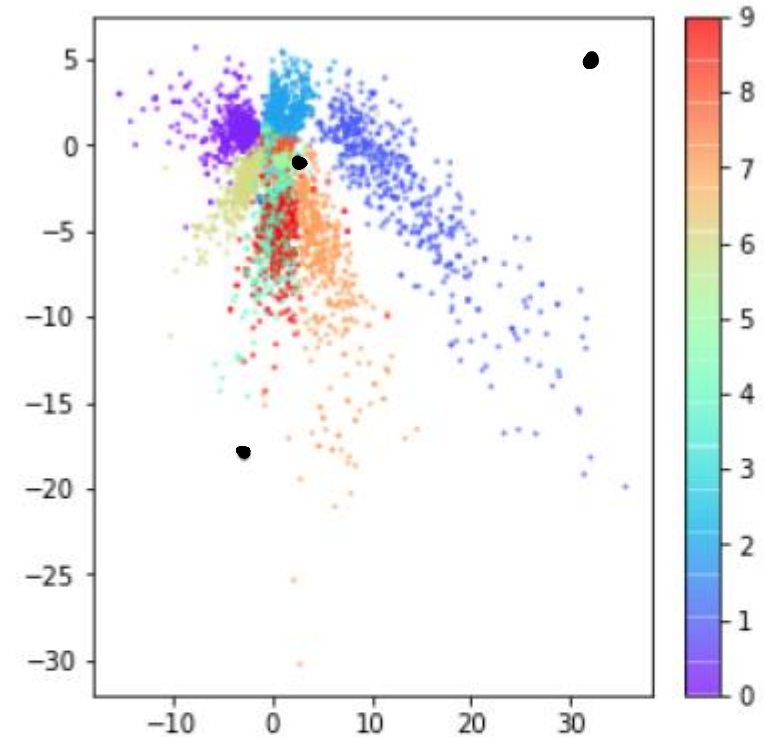
Bierzemy pod uwagę tylko różnicę między rekonstrukcją obrazu a oryginalnym obrazem. Ponadto używając MSE traktujemy wszystkie piksele z taką samą wagą, a w rzeczywistości nie wszystkie piksele powinny być traktowane równo np.: najważniejsze zazwyczaj są krawędzie. Skutkiem będzie rozmyty obraz wynikowy.

Autokoder - problemy

Chcielibyśmy, aby wybierając dowolny punkt z przestrzeni ukrytej, za pomocą dekodera generować nowy obraz.

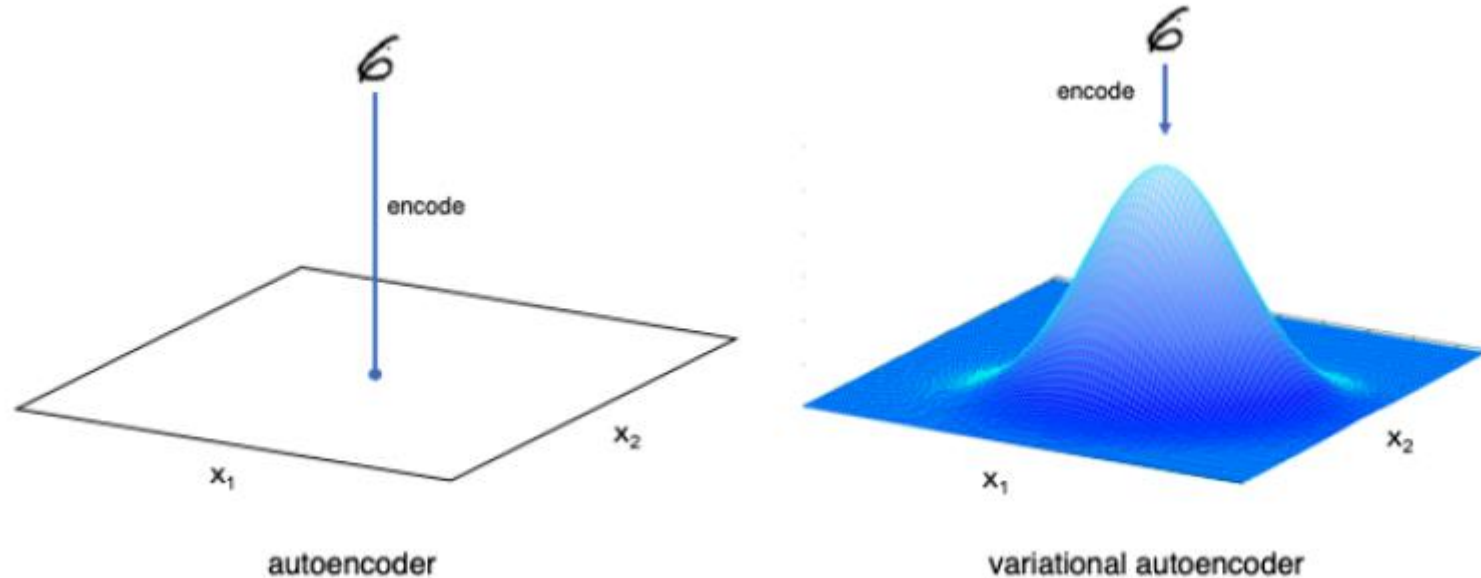
Jednak czy to w tym przypadku takie proste?

1. Dużo dziur w przestrzeni ukrytej.
2. Klasy zajmują obszary o znacznie różniące się powierzchni w przestrzeni ukrytej.

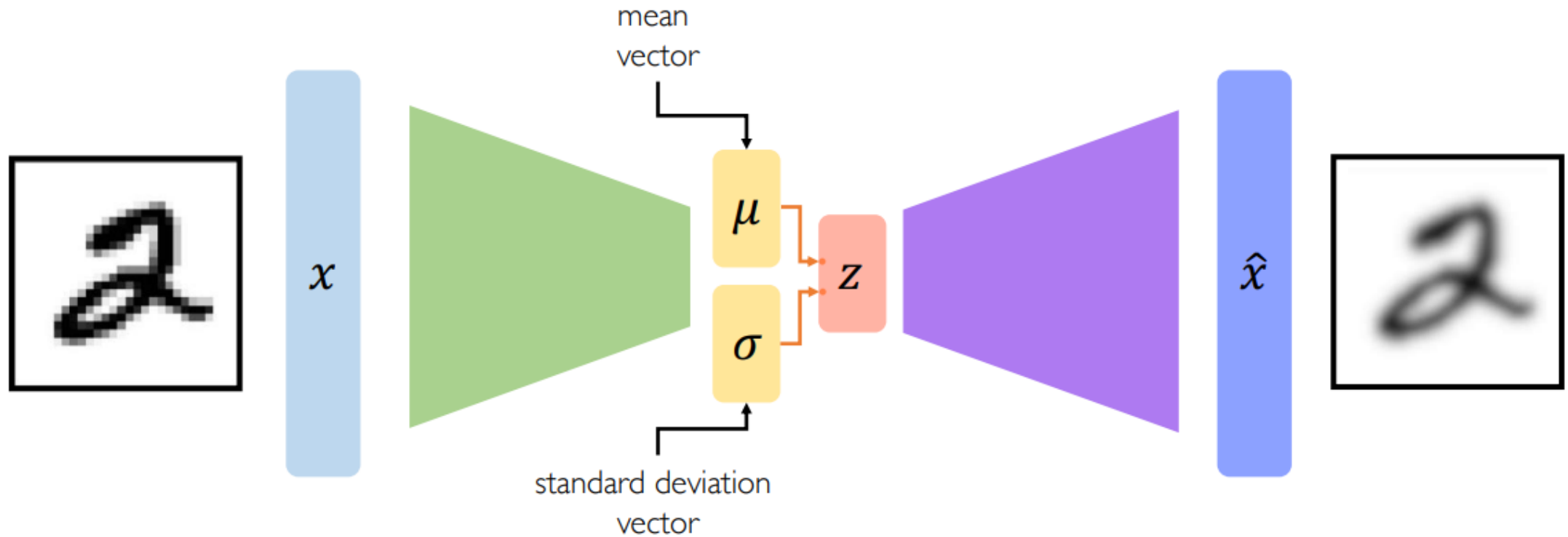


VAE: Autokoder wariacyjny

Zamiast mapować każde zdjęcie na jeden punkt w przestrzeni ukrytej, mapujemy zdjęcie na wielowymiarowy rozkład normalny wokół punktu w przestrzeni ukrytej.



VAE



VAE – funkcja straty

$$L_{VAE}(\theta, \phi) = -\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) + D_{KL} \left(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel \mathcal{N}(0, I) \right)$$

$$D_{KL}(p \parallel q) = \int_x p(x) \log \frac{p(x)}{q(x)} dx$$

p – dekodery

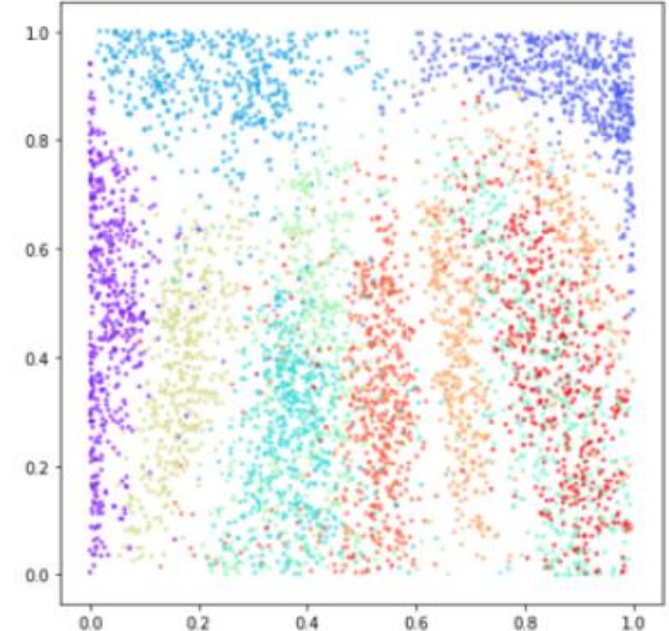
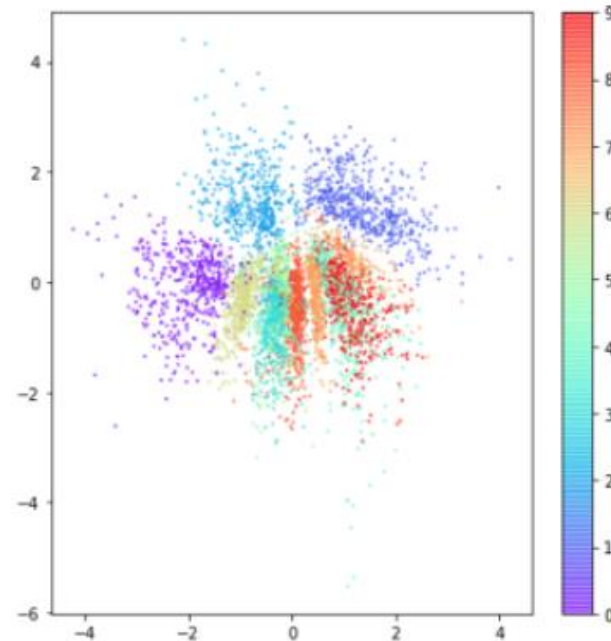
q – koder

Cel:

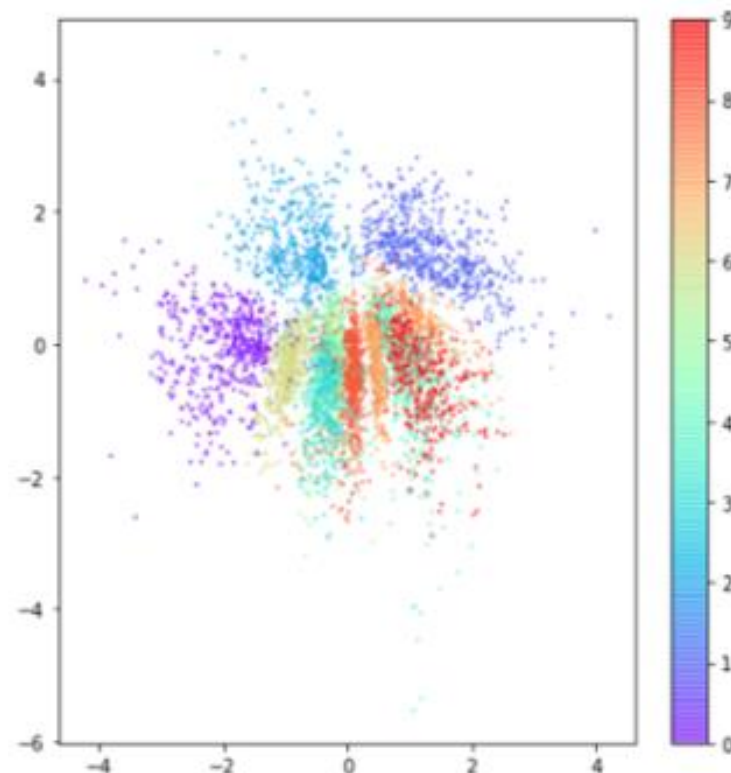
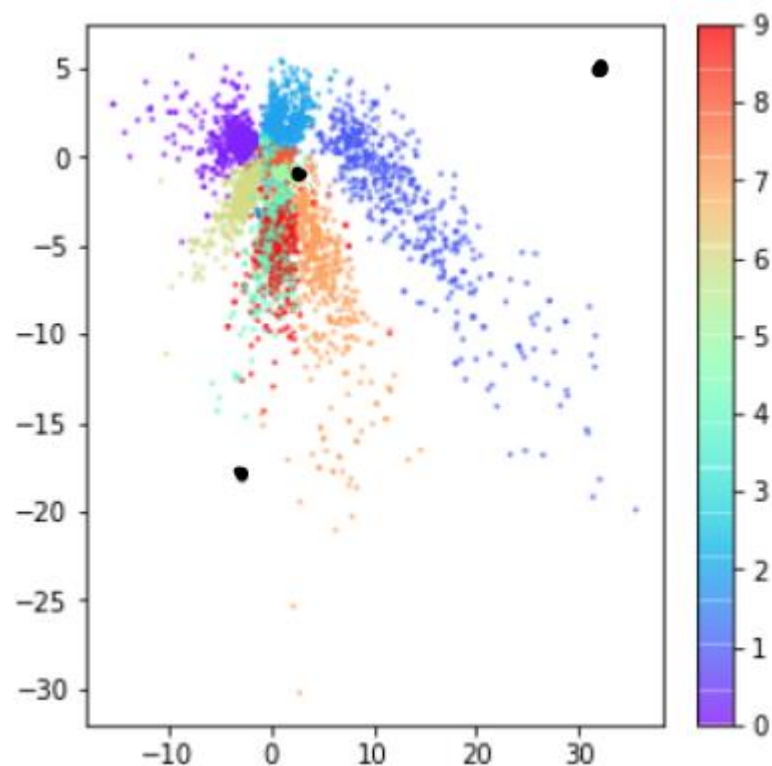
1. Wszystkie punkty położone wokół jednej wartości średniej w przestrzeni ukrytej, po przejściu przez dekodery, produkują bardzo podobne do wejścia wyjście – **reconstruction loss**.
2. Wszystkie rozkłady w przestrzeni ukrytej bardzo podobne do rozkładu normalnego – **Dywergencja Kullbacka-Leiblera**.

VAE – rozwiązane problemy

- Dobrze zdefiniowany rozkład punktów.
- Wszystkie rozkłady mają być bliskie normalnemu – małe dziury w przestrzeni.

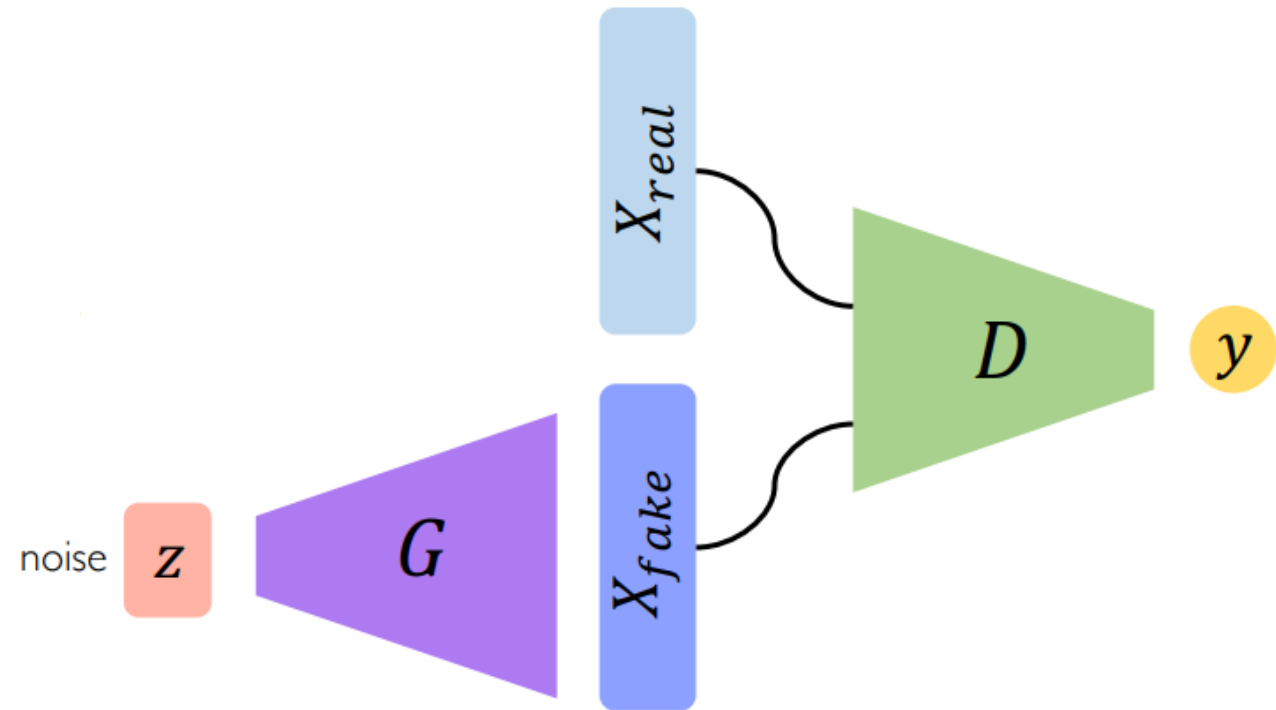


Przestrzeń ukryta – Autokoder vs VAE

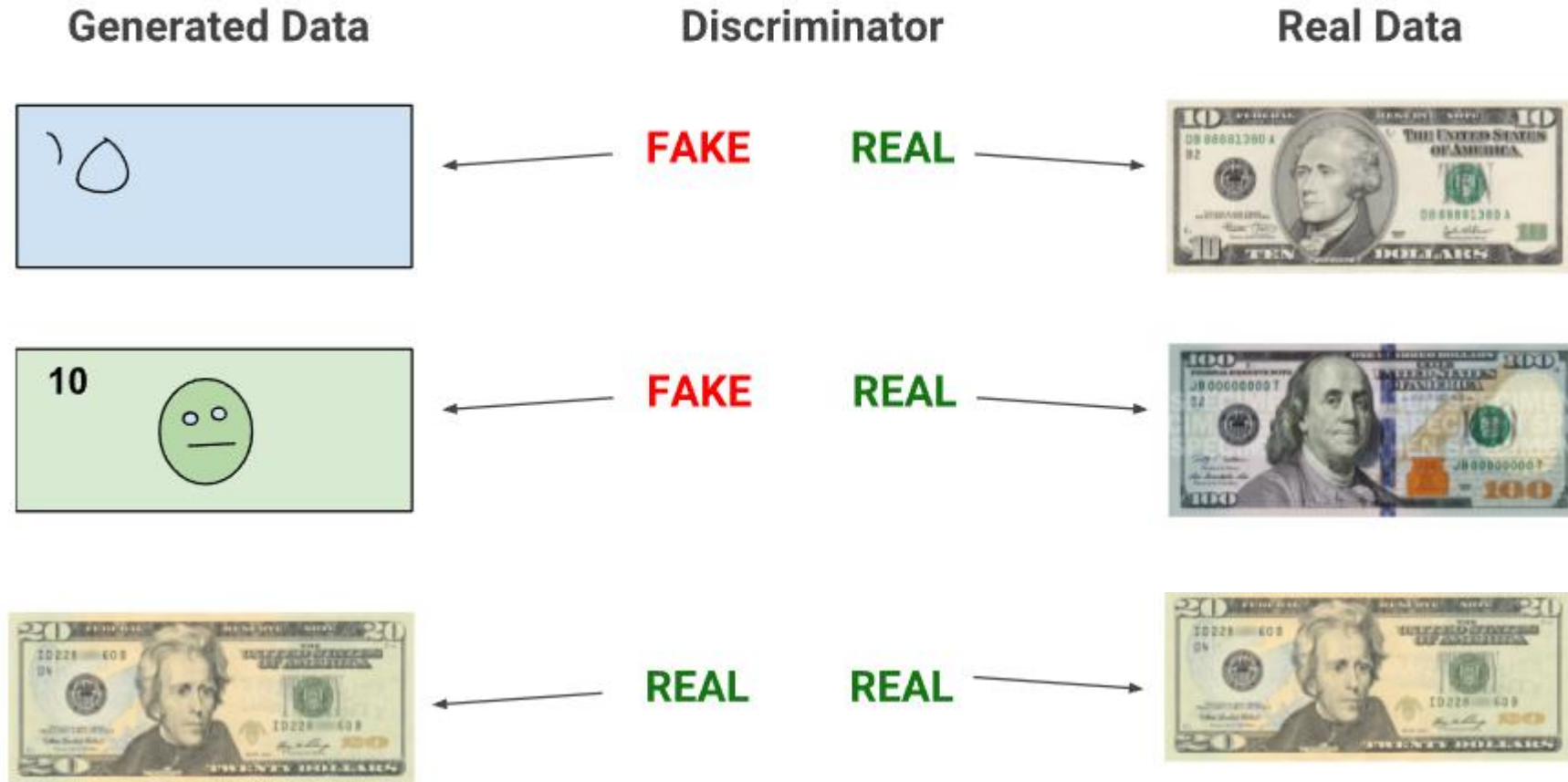


GAN – Generative Adversarial Network

- **Generator** – Stara się przekształcić losowy szum na obraz, który wygląda jak obraz ze zbioru treningowego. Wygenerowany obraz trafia na wejście dyskryminatora.
- **Dyskryminator** – Stara się odróżnić obrazy wygenerowane przez generator od prawdziwych.



GAN – Trening



GAN – Trening

$$\min_{\theta} \max_{\phi} [\mathbb{E}_{x \sim p_{data}} \log D_{\phi}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\phi}(G_{\theta}(z)))]$$

D – dyskryminator

G – generator

Dyskryminator stara się **maksymalizować** wskaźnik jakości, po to aby $D(x)$ było bliskie 1, a $D(G(x))$ było bliskie 0.

Generator stara się **minimalizować** wskaźnik jakości, po to aby $D(G(z))$ było bliskie 1.

VAEGAN

Problem: Porównywanie piksel po pikselu nie jest dobrą metryką oceniającą dla obrazów.

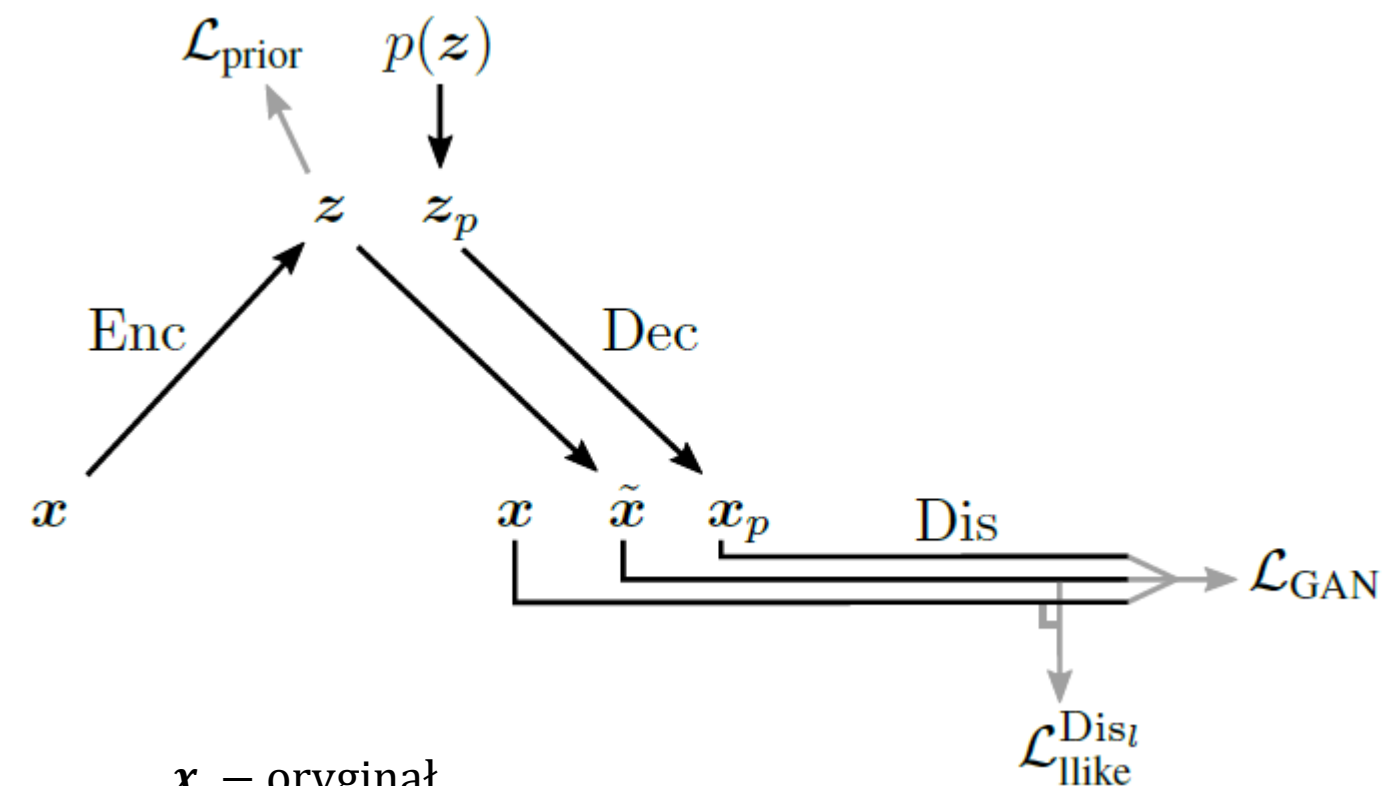
Idea: Wykorzystanie reprezentacji nauczonych przez dyskryminator w funkcji straty jako „reconstruction error”.

$$-\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(\mathbf{x}|\mathbf{z}) \quad \longrightarrow \quad -\mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \log p_{\theta}(D_{\mathbf{v}}^l(\mathbf{x})|\mathbf{z})$$

$$p_{\theta}(D_{\mathbf{v}}^l(\mathbf{x})|\mathbf{z}) = \mathcal{N}(D_{\mathbf{v}}^l(\mathbf{x})|D_{\mathbf{v}}^l(\tilde{\mathbf{x}}), I)$$

$D_{\mathbf{v}}^l$ – ukryta reprezentacja l – tej warstwy dyskryminatora

VAEGAN - Trening



x – oryginal
 \tilde{x} – rekonstrukcja
 x_p – generacja

Algorithm 1 Training the VAE/GAN model

$\theta_{\text{Enc}}, \theta_{\text{Dec}}, \theta_{\text{Dis}} \leftarrow$ initialize network parameters

repeat

$X \leftarrow$ random mini-batch from dataset

$Z \leftarrow \text{Enc}(X)$

$\mathcal{L}_{\text{prior}} \leftarrow D_{\text{KL}}(q(Z|X) \| p(Z))$

$\tilde{X} \leftarrow \text{Dec}(Z)$

$\mathcal{L}_{\text{llike}}^{\text{Disl}} \leftarrow -\mathbb{E}_{q(Z|X)} [p(\text{Disl}(X)|Z)]$

$Z_p \leftarrow$ samples from prior $\mathcal{N}(\mathbf{0}, I)$

$X_p \leftarrow \text{Dec}(Z_p)$

$\mathcal{L}_{\text{GAN}} \leftarrow \log(\text{Dis}(X)) + \log(1 - \text{Dis}(\tilde{X}))$
 $\quad + \log(1 - \text{Dis}(X_p))$

// Update parameters according to gradients

$\theta_{\text{Enc}} \xleftarrow{+} -\nabla_{\theta_{\text{Enc}}} (\mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{llike}}^{\text{Disl}})$

$\theta_{\text{Dec}} \xleftarrow{+} -\nabla_{\theta_{\text{Dec}}} (\gamma \mathcal{L}_{\text{llike}}^{\text{Disl}} - \mathcal{L}_{\text{GAN}})$

$\theta_{\text{Dis}} \xleftarrow{+} -\nabla_{\theta_{\text{Dis}}} \mathcal{L}_{\text{GAN}}$

until deadline

VAEGAN - wyniki

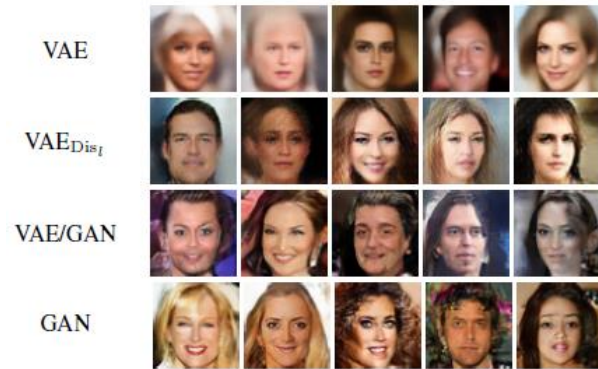


Figure 3. Samples from different generative models.



Figure 4. Reconstructions from different autoencoders.



Figure 5. Using the VAE/GAN model to reconstruct dataset samples with visual attribute vectors added to their latent representations.

Dzięki!

Jeszcze kodzik :)

Zaraz piwko 