

Jak działa Stable Diffusion?

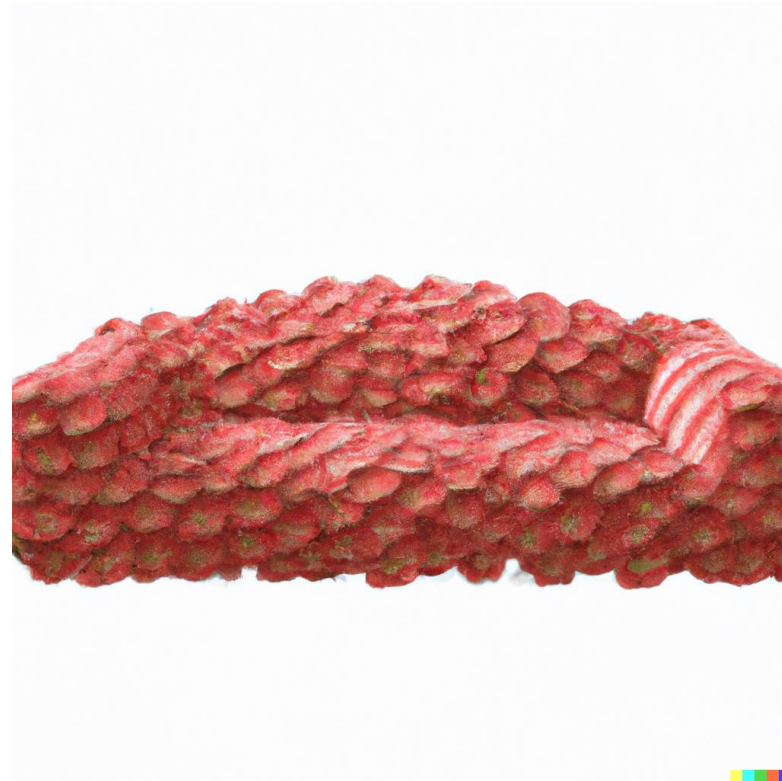
Weronika Piotrowska

sofa made of flower, simple background, design, hybrid, realistic

Stable Diffusion



DALLE-2



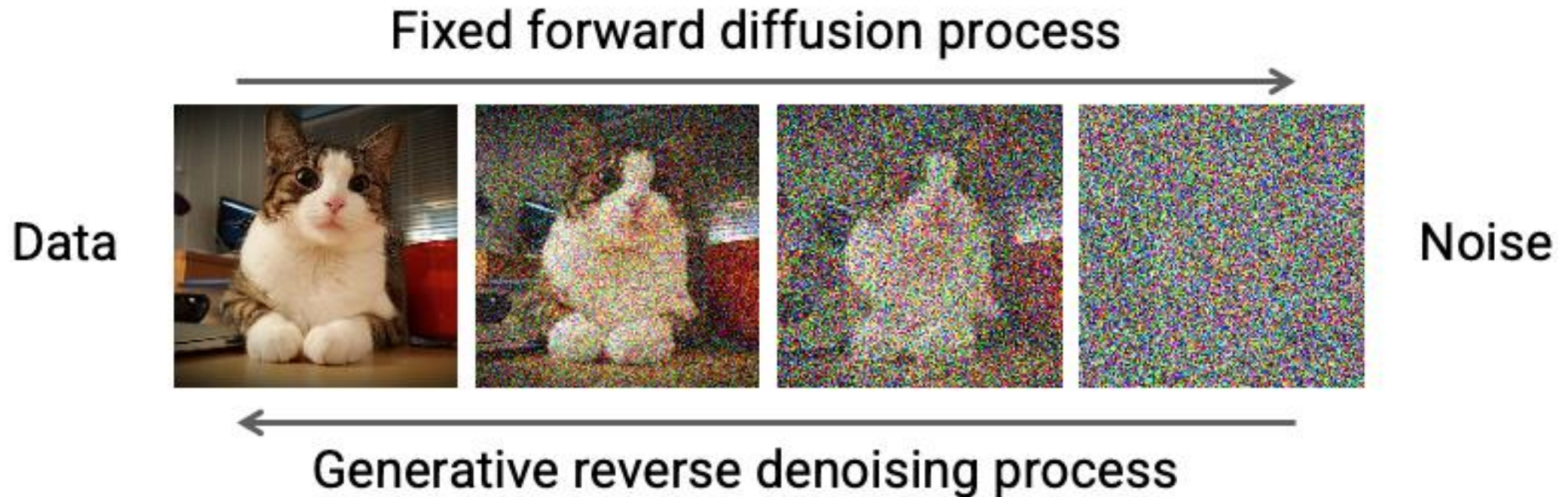
Midjourney



Agenda

1. Model dyfuzyjny
 1. Forward diffusion
 2. Odszumianie
 3. Log-likelihood
2. U-Net
3. Autoencoder
4. CLIP – tekst + obraz
- 5. Fun Quiz**
6. Kontrowersje

Diffusion models



Forward diffusion

$$q(x_t|x_{t-1}) := \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t\mathbf{I})$$

β_t - współczynnik ilości szumu dodana w chwili t

I – macierz jednostkowa

Odszumianie

Rozkład szumu dla x_{t-1} wiedząc jakie jest x_t

$$p_{\theta}(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_{\theta}(x_t, t), \Sigma_{\theta}(x_t, t))$$

Przewidujemy μ_{θ}

Σ_{θ} może być zdefiniowana jako stała

Log-likelihood

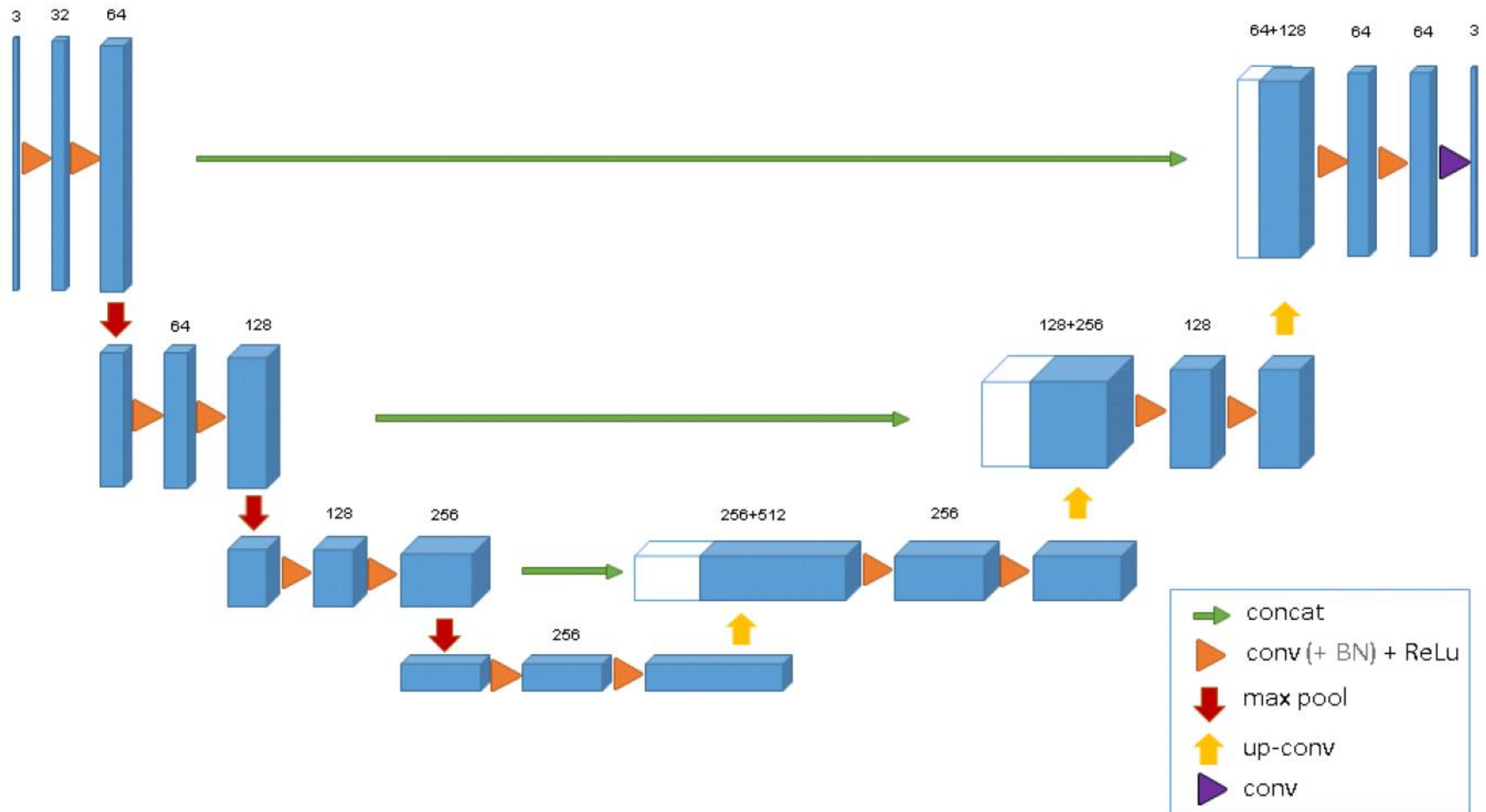
Funkcja straty

$$L_{\text{simple}} := E_{t \sim [1, T], x_0 \sim q(x_0), \epsilon \sim \mathcal{N}(0, \mathbf{I})} [\|\epsilon - \epsilon_{\theta}(x_t, t)\|^2]$$

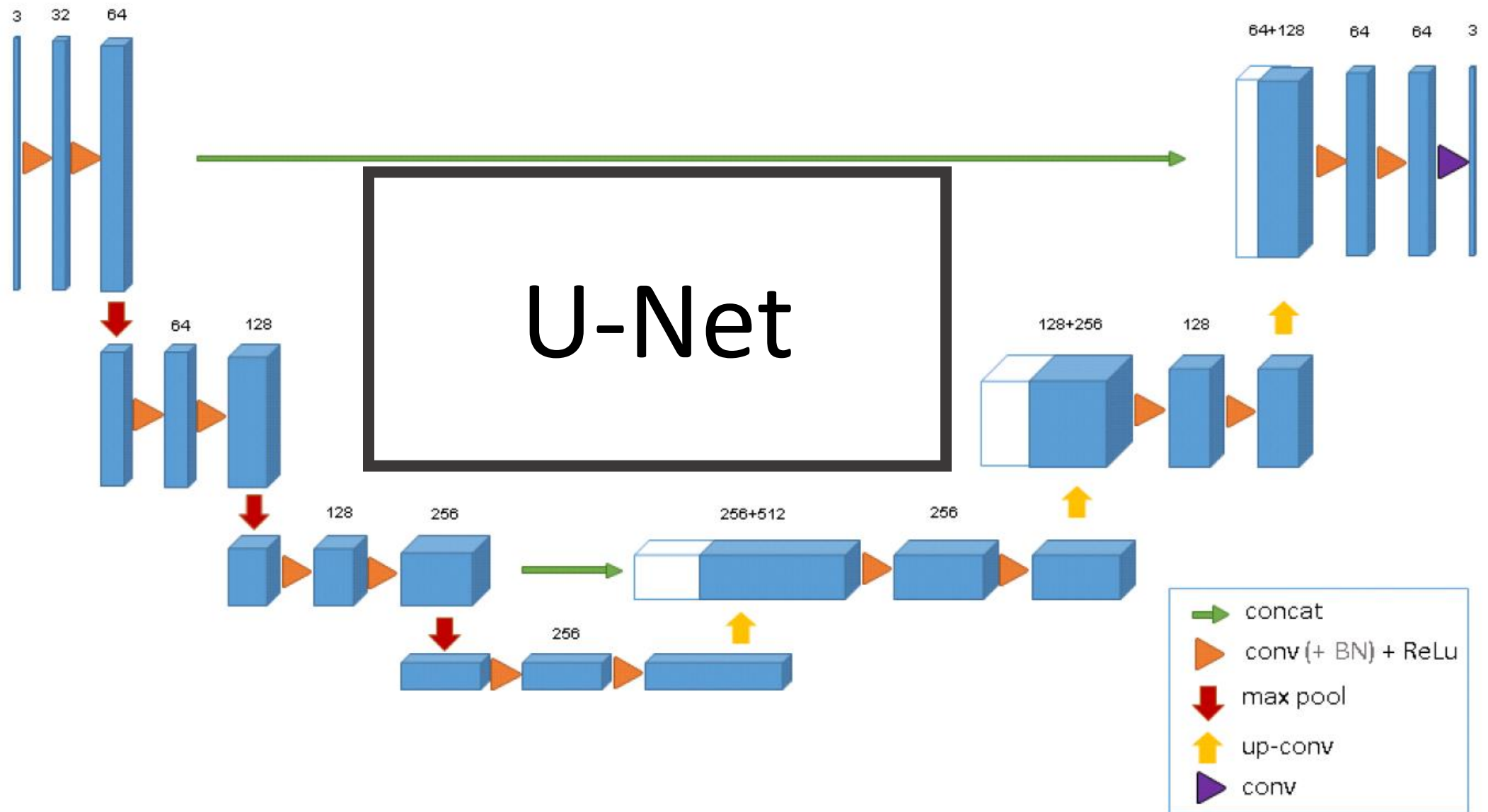
ϵ – prawdziwy szum

ϵ_{θ} - przewidywany szum

Jak nauczyć sieć odszumiania?

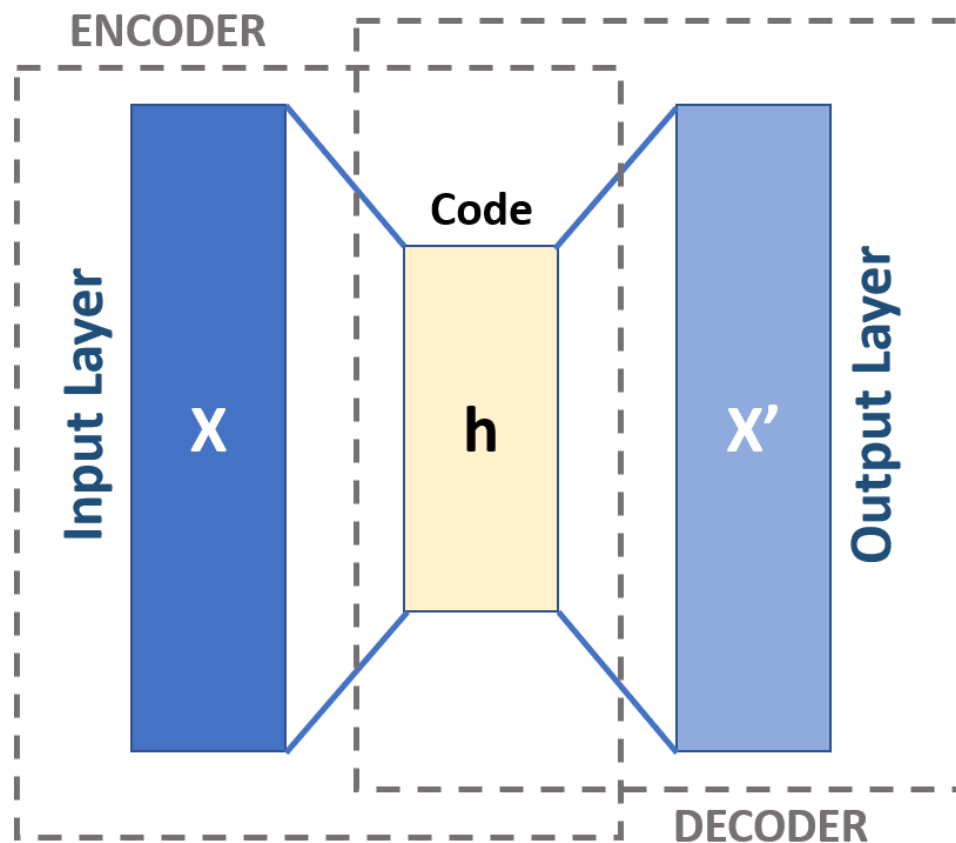


Jak nauczyć sieć odszumiania?

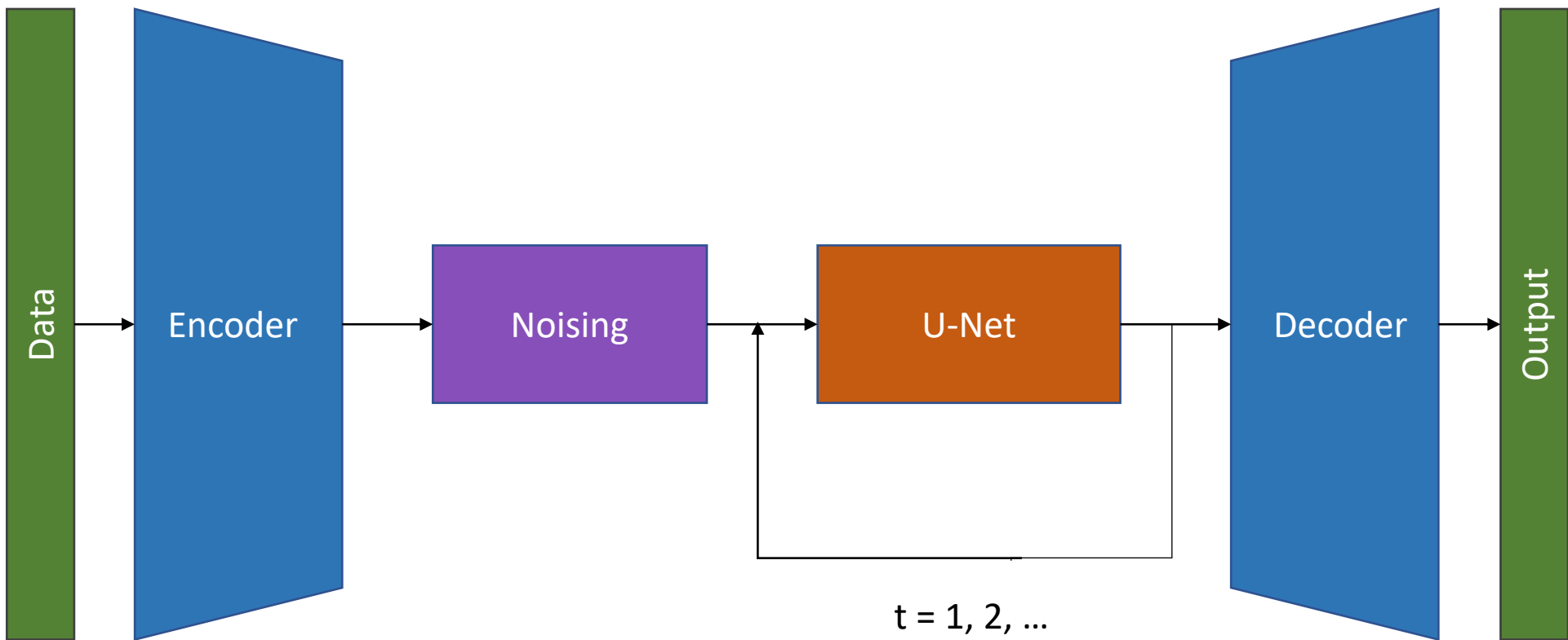


Co zrobić z odsumioną przestrzenią?

Wsadzić wszystko poprzednie w autoencoder (yay!)



So far so good



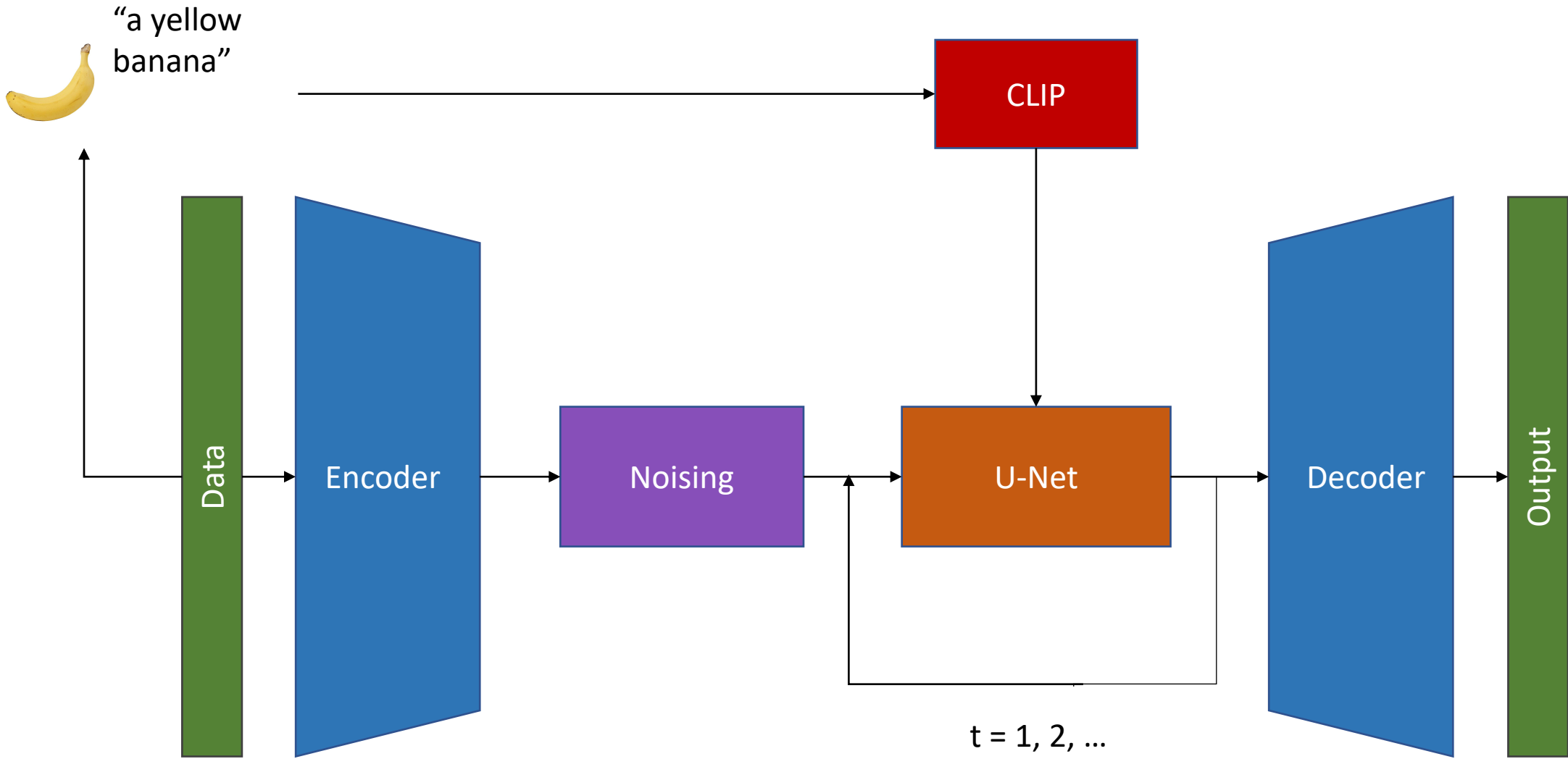
Generowanie obrazów z tekstu

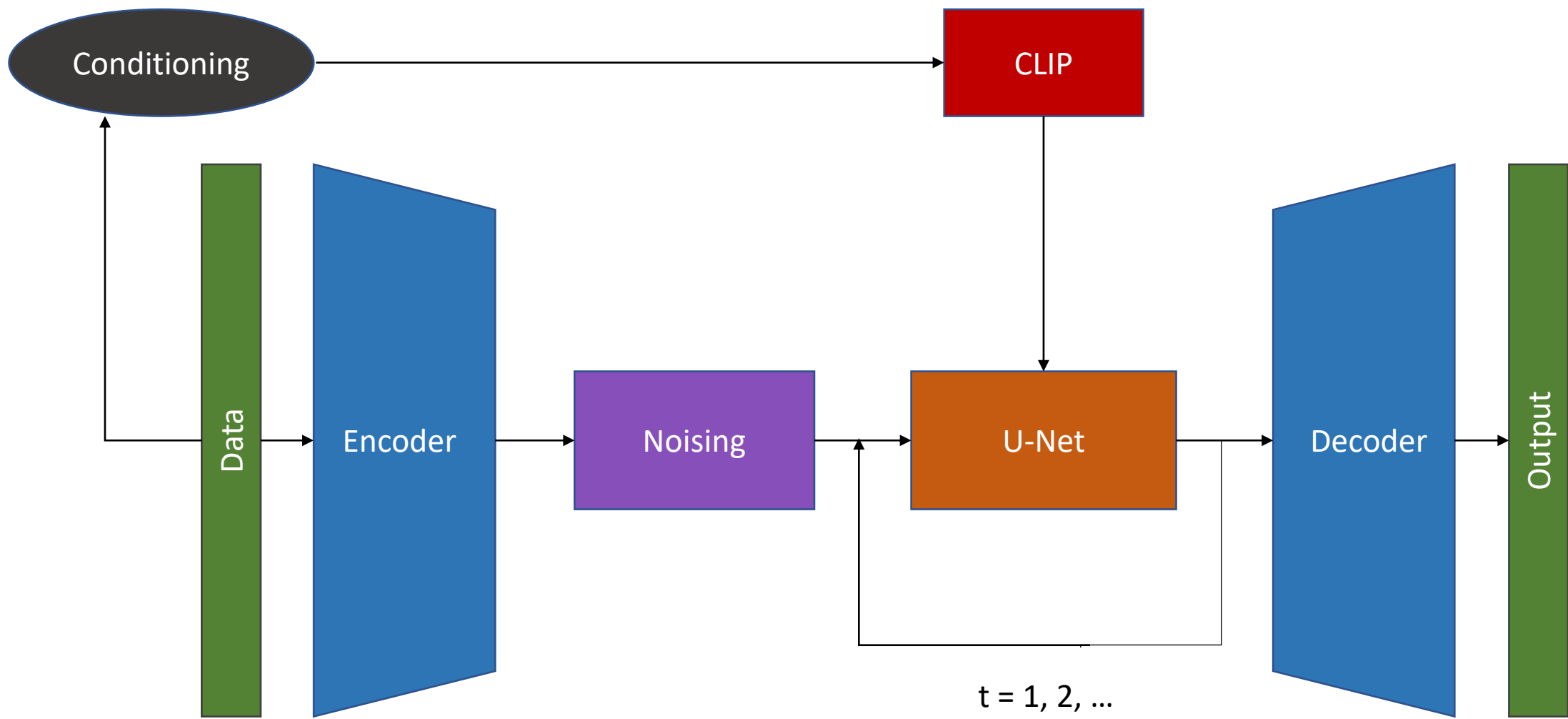
Trzeba połączyć obraz i tekst w pary - CLIP

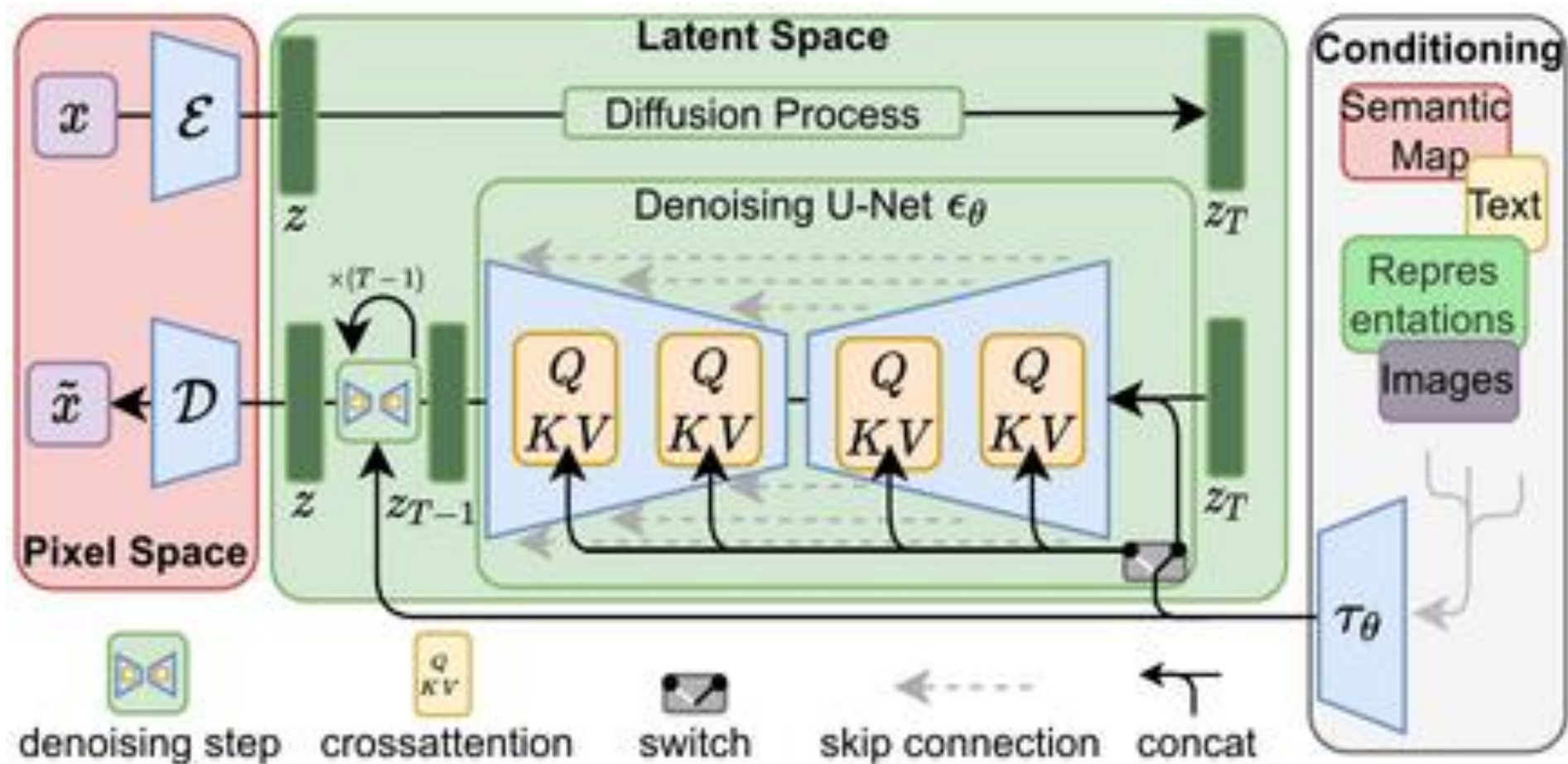


=

“banana”



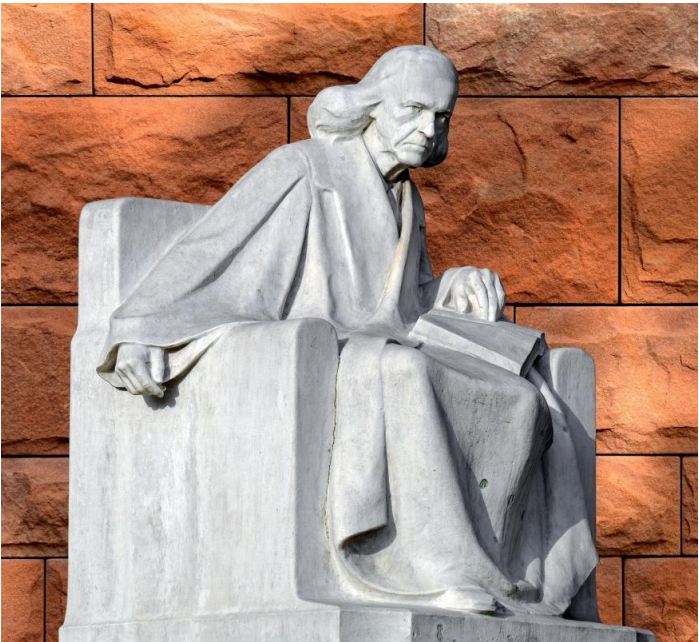


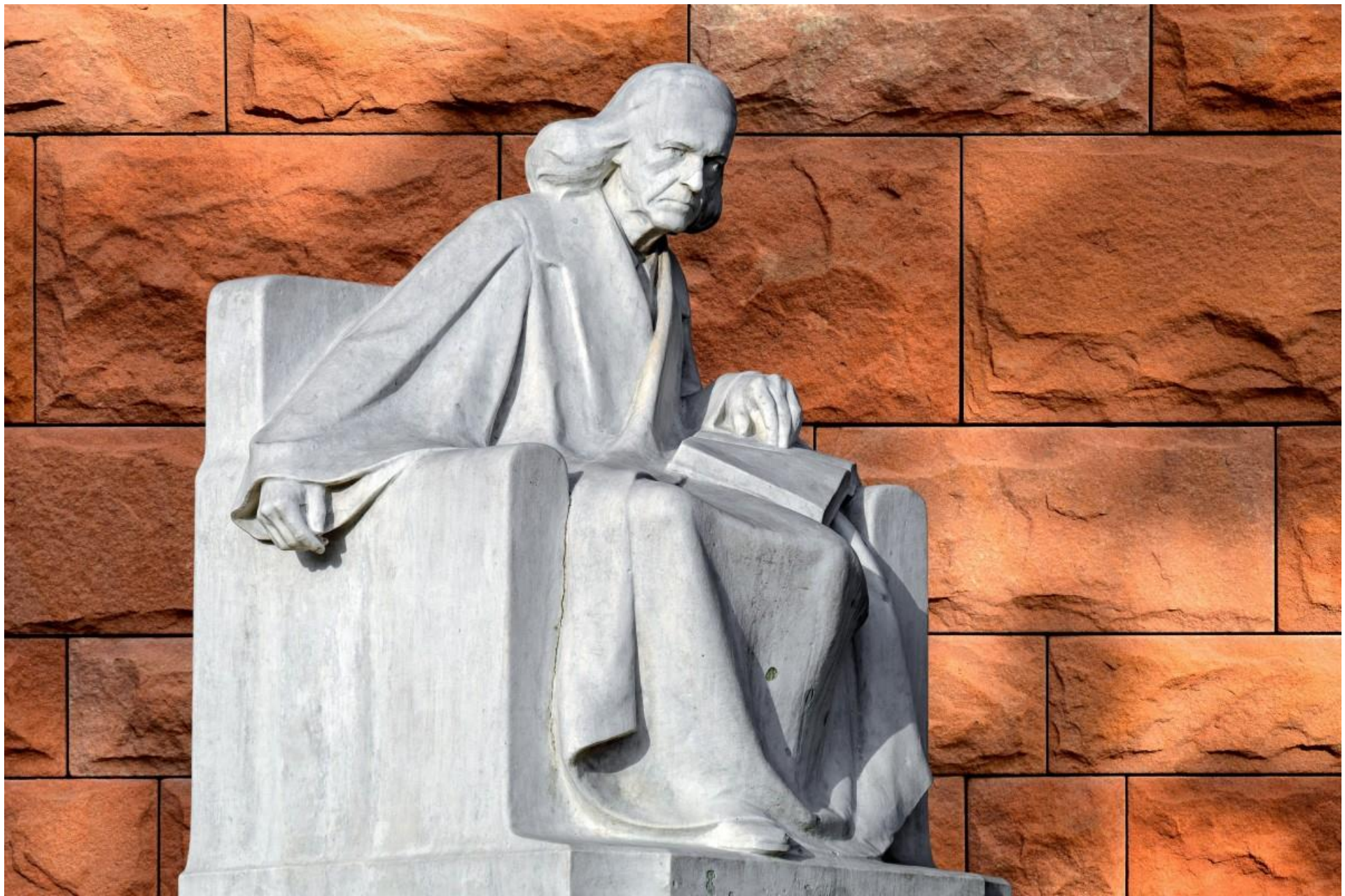


Fun Quiz!













Kontrowersje

Dzięki za uwagę



References

<https://arxiv.org/pdf/2105.05233.pdf>

<https://jalammar.github.io/illustrated-stable-diffusion/>

<https://arxiv.org/pdf/2103.00020.pdf>

<https://www.youtube.com/watch?v=W-O7AZNzbzQ>