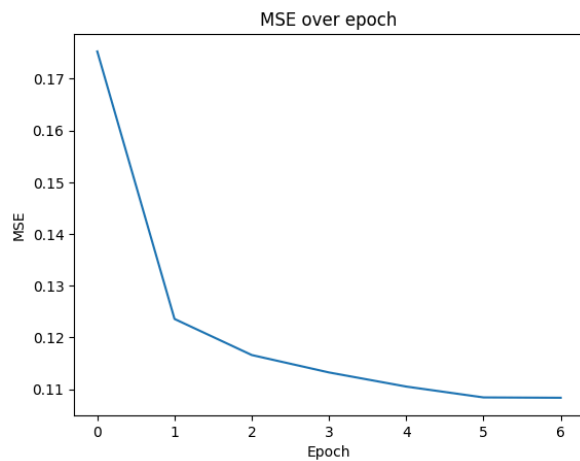


# CelebA - Stable Diffusion



Example training run



"Young blond actress with serious face" 30x  
2.0 CFG



"woman with long dark hair smiling" 30x 1.5  
CFG



"Handsome young man in tuxedo" 30x 1.5  
CFG

Fréchet inception distance: **0.123**

Inception score: **0.23 + 0.1**

Czas generacji w 30 krokach dla batcha 4 zdjęcia 256×256 (rtx 3060 12GB): ~20s

## Dlaczego Stable Diffusion?

- Jeden dobrze wytrenowany model i mamy: text to image, image to image, inpainting...
- Iteracyjny proces, przez co możemy w zależności od potrzeb balansować między jakością a szybkością generacji
- Warto zkompresować dane wejściowe i pracować mniejszych danych, dlatego używamy VAE: oszczędzamy zasoby, zachowujemy "płynną" przestrzeń ukrytą.
- CLIP + CFG conditioning: unikamy zbędnego treningu dopasowania zdjęcie-text, a metoda CFG pozwala nam sterować jak silny wpływ powinien mieć prompt

## Dane wejściowe

CelebA dataset: <https://www.kaggle.com/datasets/jessicali9530/celeba-dataset>

Dane dzielone losowo:

- 60% UNet
- 40% VAE

A następnie w każdym komponencie

- 70% train
- 20% validation
- 10% test

## Analiza końcowa

Osiągnięte rezultaty są obiecujące, kontynuacja treningu + ewentualne dodatkowe augmentacje danych lub zwiększenie liczby parametrów mogłyby pomóc

poprawić aktualne wyniki.