

# Fino podešavanje diffusion modela:

## Stable Diffusion LoRA za generisanje sprske narodne nošnje

student: Tina Mihajlović SV3/2020

asistent: Marko Njegomir

# O Diffusion modelima

*“Experience without theory is blind...”*

# O Diffusion modelima

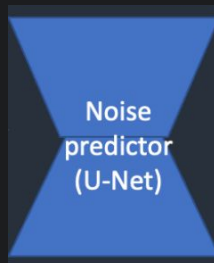
- Diffusion modeli su generativni modeli - znaju da generišu nove podatke (slike, audio, tekst...) na osnovu onoga što su videli tokom treniranja
- Diffusion modeli generišu slike (i videoe)

# O diffusion modelima

- Tadaaa!

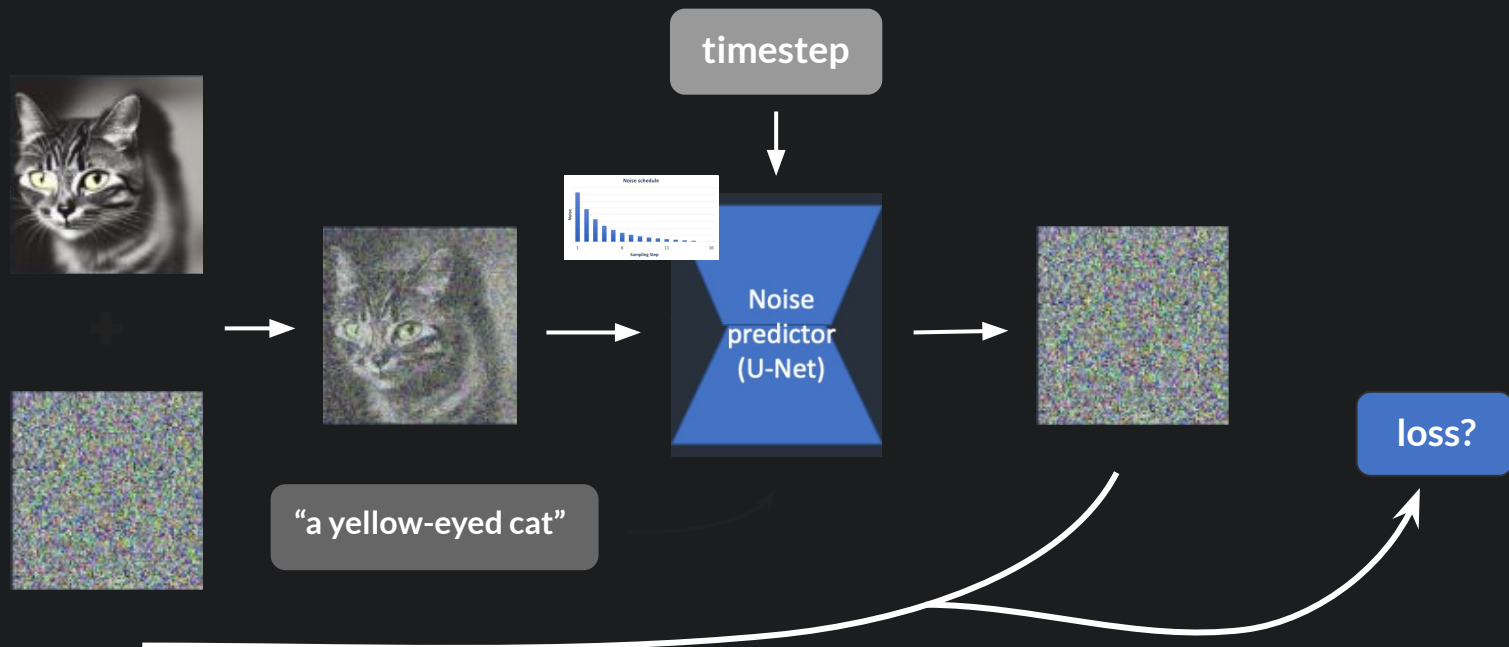


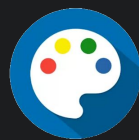
# “Magična” neuronska mreža



# “Magična” neuronska mreža

- ...jednostavno predviđa šum za dati timestep, koristeći noise schedule, uslovljena tekstualnim prompt-om





**Stable Diffusion**  
by stability.ai

# Stable Diffusion

- Latentni diffusion model za generisanje slika

“Teddy bears wearing  
suit discussing a  
business proposal  
around office table”

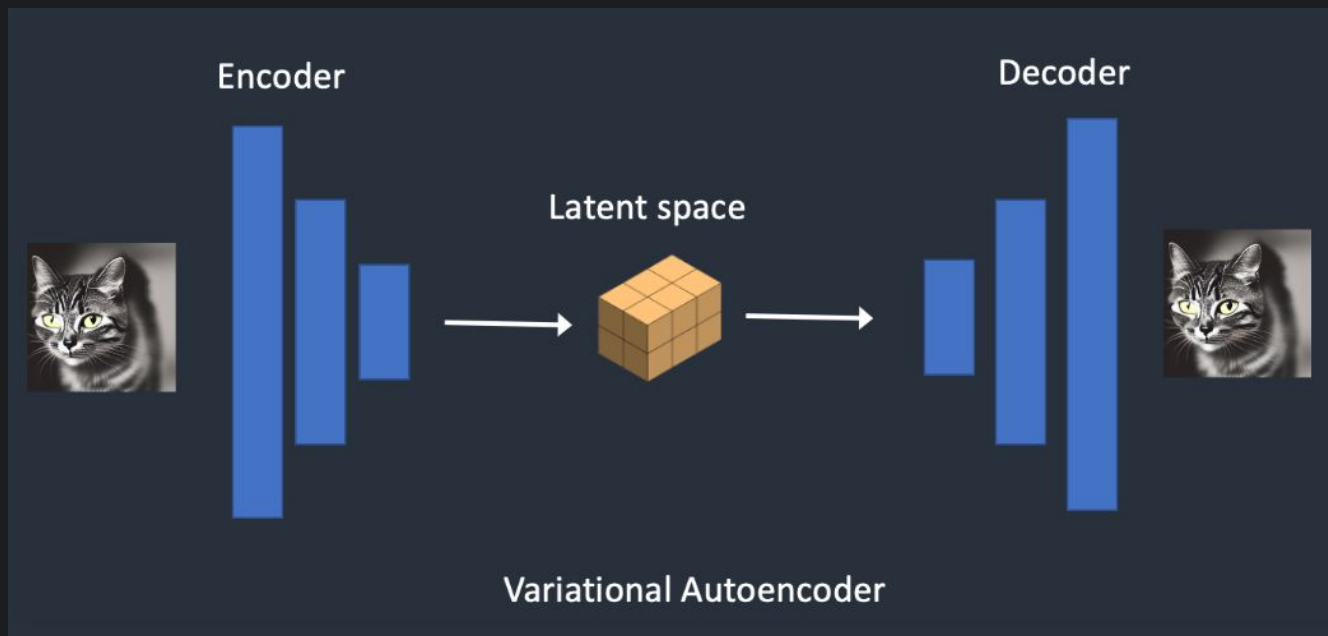
Stable  
Diffusion





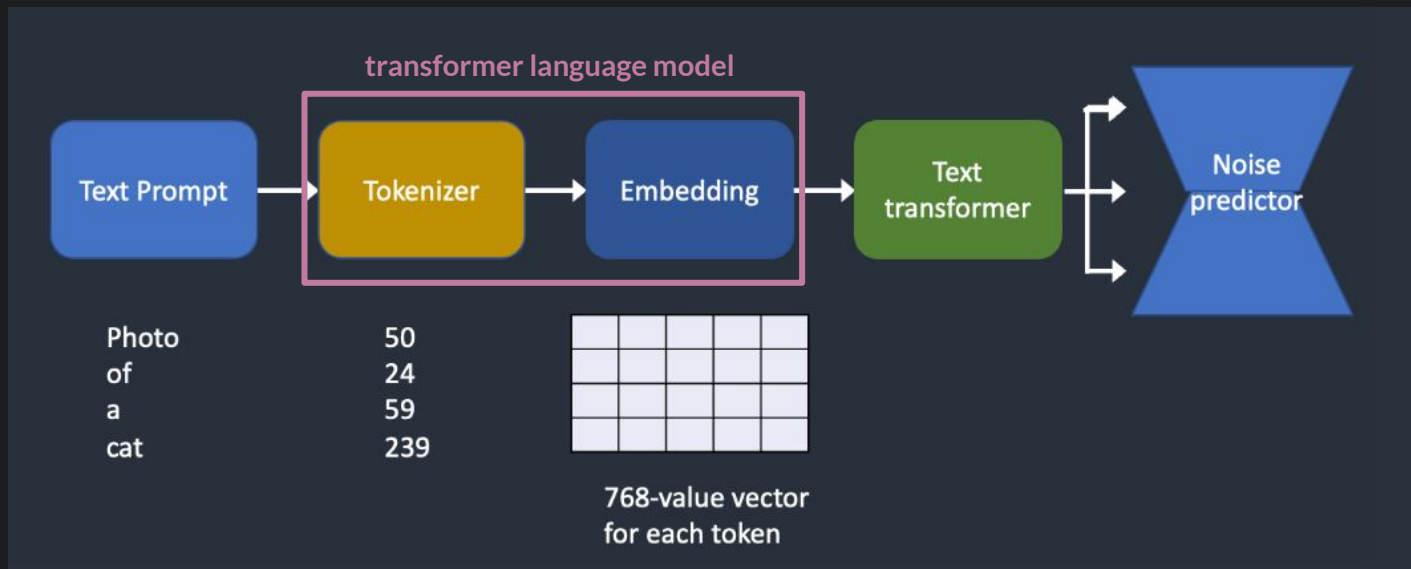
# Stable Diffusion

- Latentni diffusion model za generisanje slika
- 48x manje brojeva => brže => potreban manje zahtevan hardver



# Tekstualno uslovljavanje

- Cilj je da usmerimo noise predictor da predvidi šum koji će, kada ga oduzmemo od slike, da nam da ono što želimo da generišemo



# Tekstualno uslovljavanje - attention is all we need

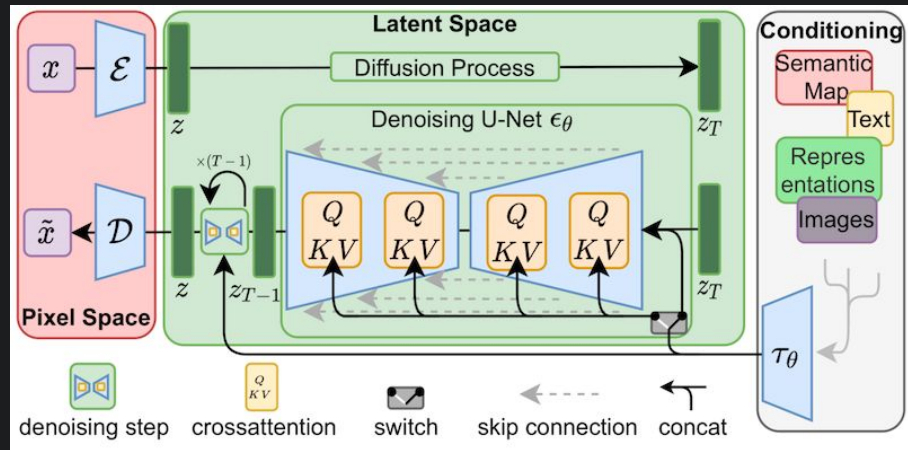
- Dodajemo attention slojeve našem Unet noise predictor
  - Self-attention - “a man with blue eyes wearing a shirt”
  - Cross-attention - semantička veza teksta i slike

# Stable Diffusion - problem sa:

Ubacivanje novih karaktera/stilova

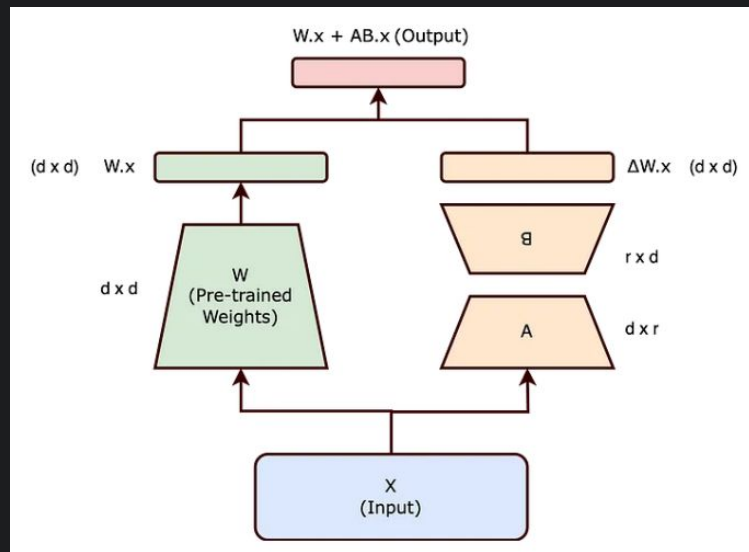
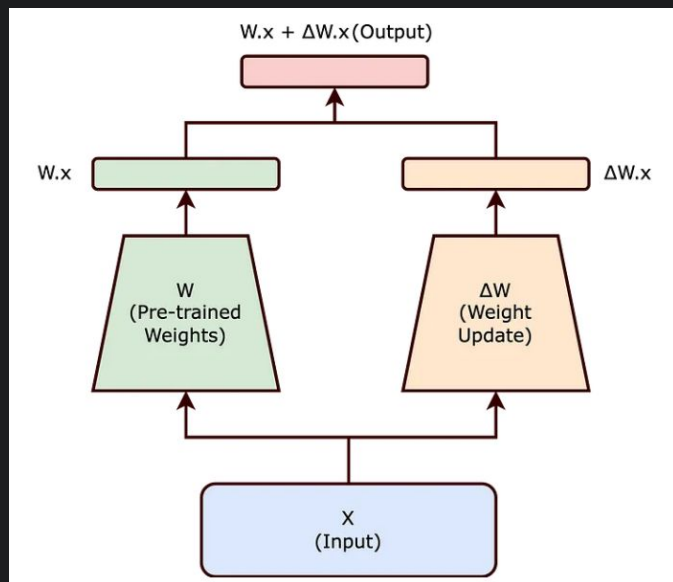
# Low Rank Adaptation (LoRA)

- “A powerful way to insert new concepts into the model while saving on space and power.”
- Cilj: naučiti model novi koncept/stil/karaktera
- Fokus na cross-attention slojevima
- Rezultat: odvojeni fajlovi koji se nadodaju na checkpoint modele
- Veličina fajlova: 2-300MBs
- Veličina dataset-a: 10+ slika



# LoRA - osnovna ideja

- Zamrzni osnovne težine, uči samo razlike kroz injektovane matrice
- Dodajemo nove težine u cross attention slojeve Unet mreže
- Low rank? ideja
- *Na čemu smo sve uštedeli?*



# LoRA - primena na projektu

- Zadatak: naučiti model da generiše srpsku narodnu nošnju



+ labele (BLIP model + best practice)

# Priprema skupa podataka



## **BLIP:**

"a girl in a folk costume standing on a set of stairs"

## **Trigger reč + best practice:**

"nosnjaoutfit, a girl in a traditional nosnjaoutfit standing on a set of stairs, wood fence, green trees"



# Ključna (trigger) reč



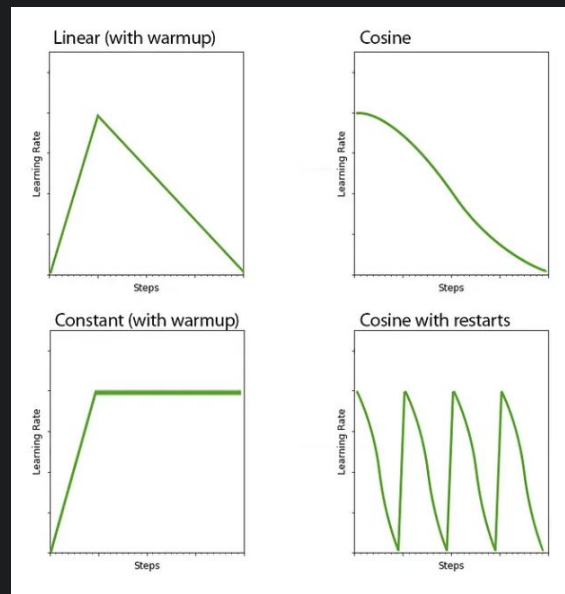
bez



sa

# Treniranje

- Izbor “baznog modela” - Stable Diffusion v1.5
  - general purpose
- Igranje sa hiperparametrima:
  - Learning rate (1e-5, 1e-6)
  - Lr schedulers
    - Linear
    - Cosine with restarts
  - Batch size (5-8)
  - Clip Skip (1-2)
  - Epohe - 10
  - Repeats - 10



# Rezultati

- Dve (tri) verzije
  - V1
  - V2
  - (V3)



# Rezultati

- Dve (tri) verzije
  - V1
  - V2
  - (V3)





# Analyze

- XYZ plotovi - trigger reč - “nosnjaoutfit”

pixar.safetensors [732d0dd2cf]



dreamshaper.safetensors  
[879db523c3]



anything-v3-fp16-pruned.safetensors  
[d1facd9a2b]



pixar.safetensors [732d0dd2cf]



dreamshaper.safetensors  
[879db523c3]

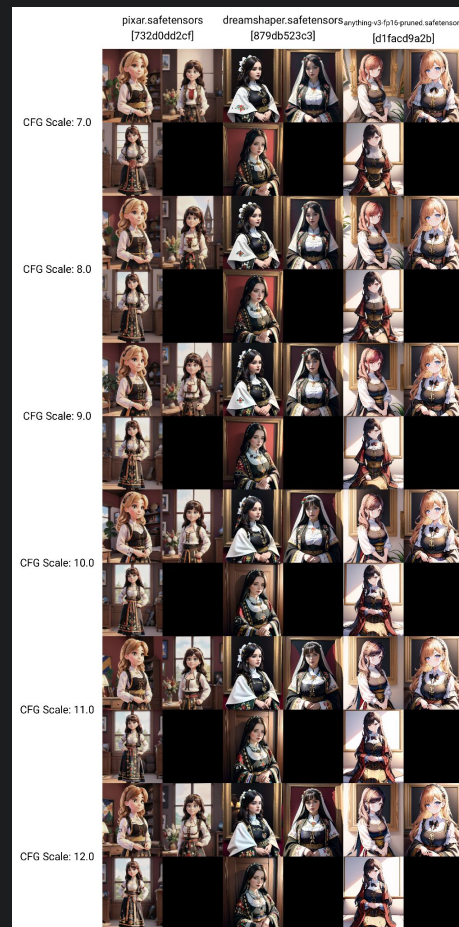


anything-v3-fp16-pruned.safetensors  
[d1facd9a2b]



# Analyze

- XYZ plotovi - uticaj parametara



# Analyze

- XYZ plotovi - poređenje rezultata

pixar.safetensors [732d0dd2cf]



dreamshaper.safetensors  
[879db523c3]



anything-v3-fp16-pruned.safetensors  
[d1facd9a2b]



pixar.safetensors [732d0dd2cf]



dreamshaper.safetensors  
[879db523c3]



anything-v3-fp16-pruned.safetensors  
[d1facd9a2b]





# Analyze

- XYZ plotovi - poređenje rezultata

pixar.safetensors [732d0dd2cf]



dreamshaper.safetensors  
[879db523c3]



anything-v3-fp16-pruned.safetensors  
[d1facd9a2b]



pixar.safetensors [732d0dd2cf]



dreamshaper.safetensors  
[879db523c3]



anything-v3-fp16-pruned.safetensors  
[d1facd9a2b]





# Zaključak i unapređenja

- 30 slika, 15 minuta treniranja -> zavidni rezultati
- Poboljšanja: veći skup podataka, regularizacija, hiperparametri
- Povremeni problem zamućenih lica?

# Hvala na pažnji!

---

Pitanja? 🙋

# Reference

---

- Git repozitorijum:
  - <https://github.com/tince250/sd-nosnja-lora-training>
- Korisni linkovi
  - <https://stable-diffusion-art.com/>
  - <https://hoshikat.hatenablog.com/entry/2023/05/26/223229#LoRA%E3%81%AE%E4%BB%95%E7%B5%84%E3%81%BF%E3%82%92%E7%9F%A5%E3%82%8D%E3%81%86>
  - <https://towardsdatascience.com/understanding-lora-low-rank-adaptation-for-finetuning-large-models-936bce1a07c6>
  - <https://jalammar.github.io/illustrated-stable-diffusion/>
  - <https://www.youtube.com/watch?v=HoKDTa5jHvg>
  - <https://huggingface.co/blog/lora>
  - <https://learn.deeplearning.ai/diffusion-models>
  - <https://arxiv.org/pdf/2006.11239.pdf>
  - <https://arxiv.org/pdf/2302.05543.pdf>
  - <https://www.youtube.com/watch?v=BePQBWPnYuE>