УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА У НОВОМ САДУ

Фино подешавање *Stable Diffusion* модела применом *LoRA* метода

Кандидат:
Тина Михајловић SV 3/2020

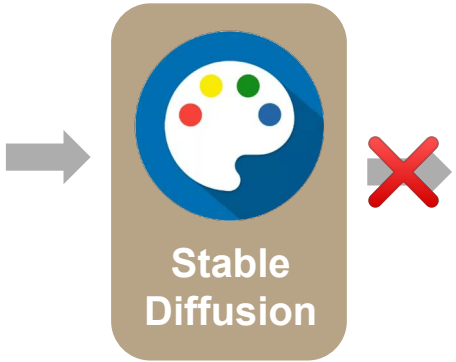
Ментор:
Проф. Др Јелена Сливка

“Фино подешавање *Stable Diffusion* модела применом *LoRA* метода”

“A girl with brown hair wearing a kimono, bamboo forest, nightime, portrait photo”

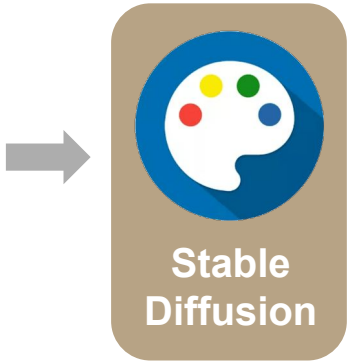
“Фино подешавање *Stable Diffusion* модела применом *LoRA* метода”

“A girl wearing traditional Serbian clothing”

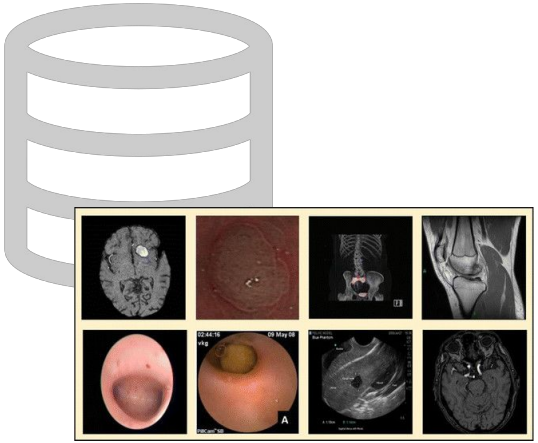


“Фино подешавање *Stable Diffusion* модела применом *LoRA* метода”

“A girl wearing traditional Serbian clothing”



Зашто се бавимо генерисањем слика?



Величина тржишта:

US$ 5201 Million

Генеративни модели

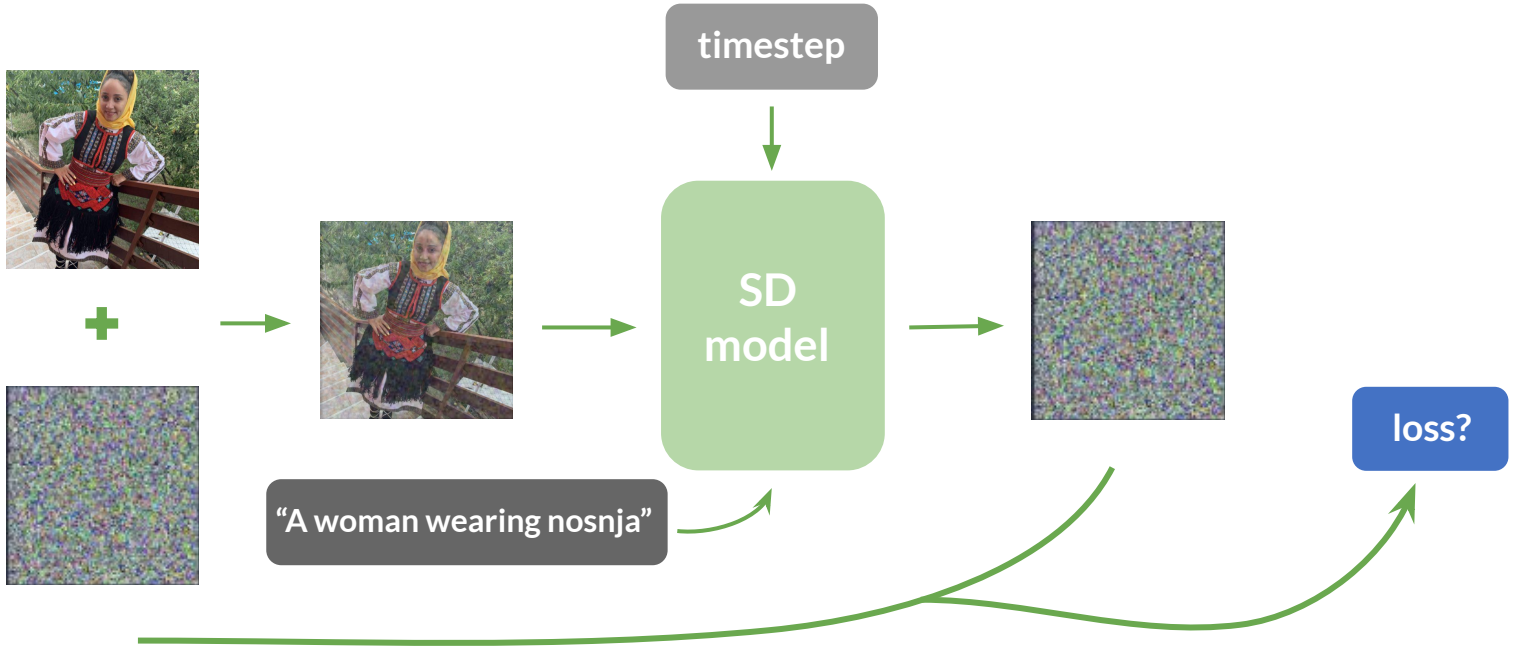
- **Генеративни модели** - знају да генеришу нове податке (слике, аудио, текст…) на основу онога што су видели током тренирања.
  - VAE, GAN, дифузни модели

Дифузни модели

- Генеративни модели (слике)
- Од потпуног шума до јасне слике
  - итеративно

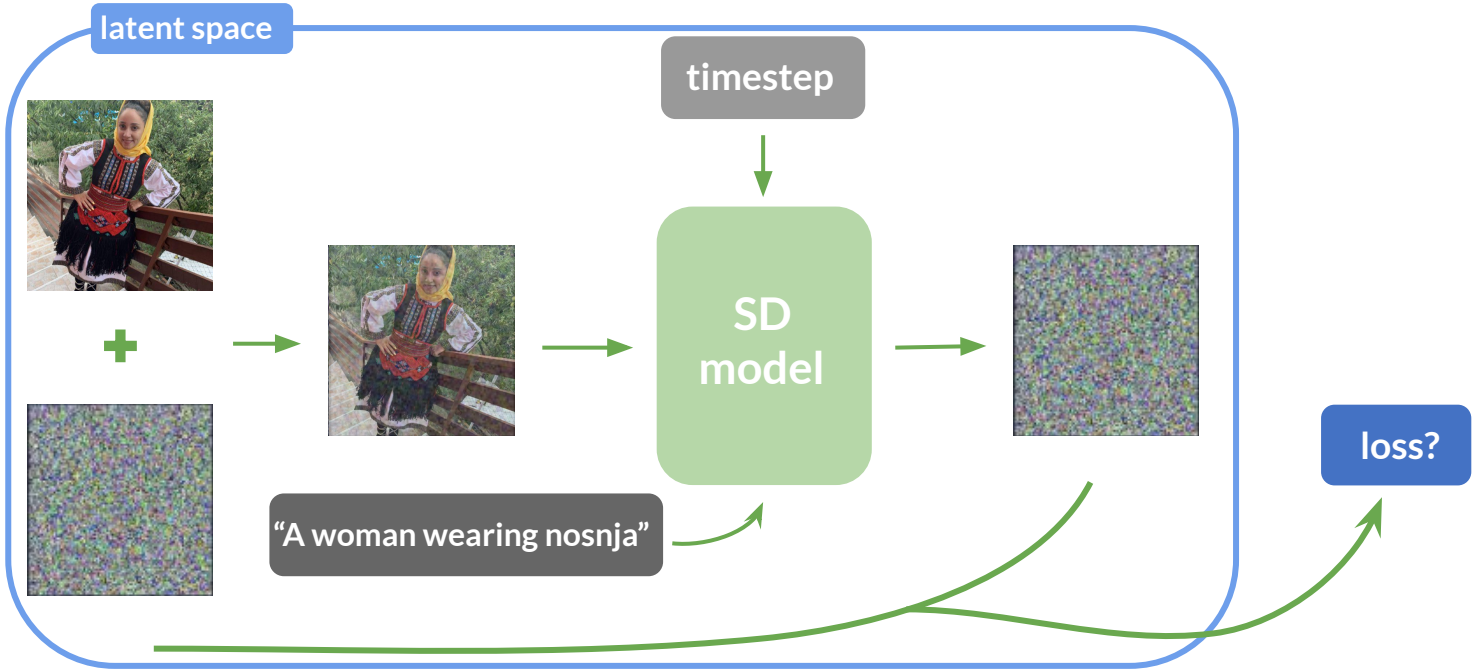
Дифузни модели

- Модел предвиђа додати шум за дати временски корак, према распореду шума, условљен текстуалним описом.



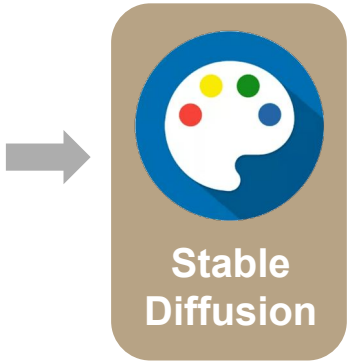
Stable Diffusion

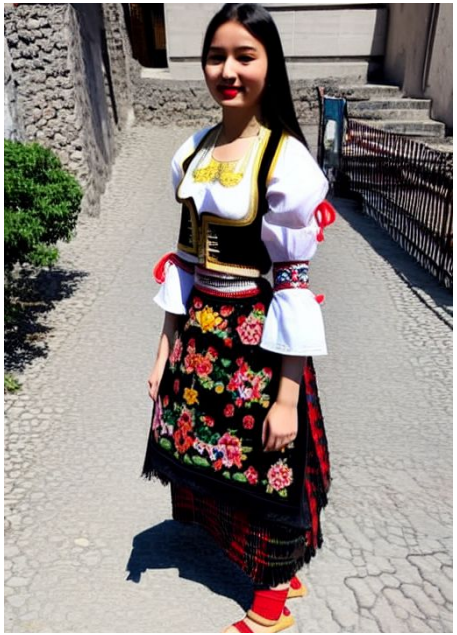
- **Латентни** дифузни модел за генерисање слика.
- 48x мање бројева => брже и мање меморије заузима



Stable Diffusion - српска ношња

“A girl wearing traditional Serbian clothing”





LoRA (Low-Rank Adaptation)

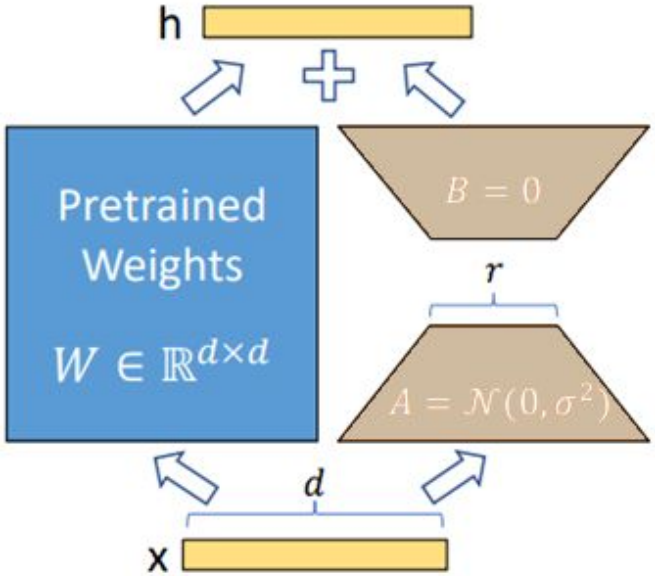
Брзо 🚀

Експресивно 👩🏾‍🎨

Меморијски ефикасно 🗄️

- Циљ: научити модел нови концепт/стил/карактера
- Матрице се инјектују у *cross-attention* слојеве

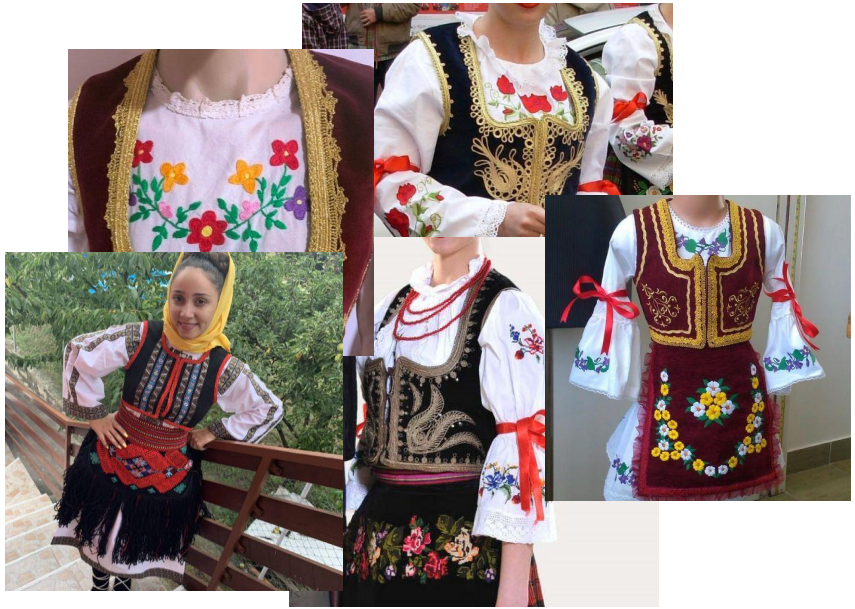
- Резултат: одвојени фајлови
- Величина фајлова: 2-300MB
- Величина скупа података: 20+ слика



LoRA - примена на пројекту

1. Прикупљање података

- Text-to-image => text + image



\+ labele

1. Прикупљање података - претпроцесирање слика

- Женска ношња
- Шумадија
- 512 x 512 px
- Исецање

1. Прикупљање података - претпроцесирање слика

- Разноврсност поза, позадина и контекста

жена у ношњи

ношња на лутки

ношња на штеклицама

горњи део ношње

доњи део ношње

жена седи у ношњи

2. Анотирање података

BLIP лабела:

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

Финална лабела:

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

3. LoRA обучавање

- Базни модел: **Stable Diffusion v1.5**
- *Kohya* скрипте
- Четири експеримента:
  - LoRA v1
  - LoRA v2
  - LoRA v3
  - LoRA v4

3. LoRA обучавање - експерименти

- Четири експеримента:
  - **LoRA v1**
    - DS_v1
    - train_batch_size: 1
    - unet_lr: 1e-4
    - text_encoder_lr: 5e-5
    - network_dim: 32
    - network_alpha: 32
  - LoRA v2
  - LoRA v3
  - LoRA v4

3. LoRA обучавање - унапређење скупа података

- Избачене слике лошег квалитета
- Проширен скуп новим сликама - 34 слике
- Лабеле додатно рафиниране

Експеримент 1 - лабела:

"nosnjaoutfit, a girl in a traditional nosnjaoutfit is posing for a picture"

Експеримент 2 - лабела:

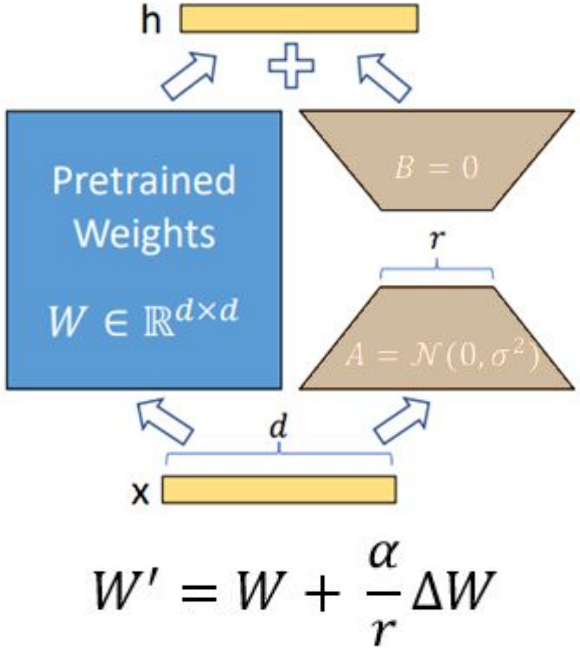
"nosnjaoutfit, a girl in a traditional nosnjaoutfit is posing for a picture, bushes in the background"

3. LoRA обучавање - експерименти

- Четири експеримента:
  - LoRA v1
  - **LoRA v2**
    - DS_v2
    - train_batch_size: 8
    - unet_lr: 1e-4
    - text_encoder_lr: 5e-5
    - network_dim: 32
    - network_alpha: 32
  - LoRA v3
  - LoRA v4

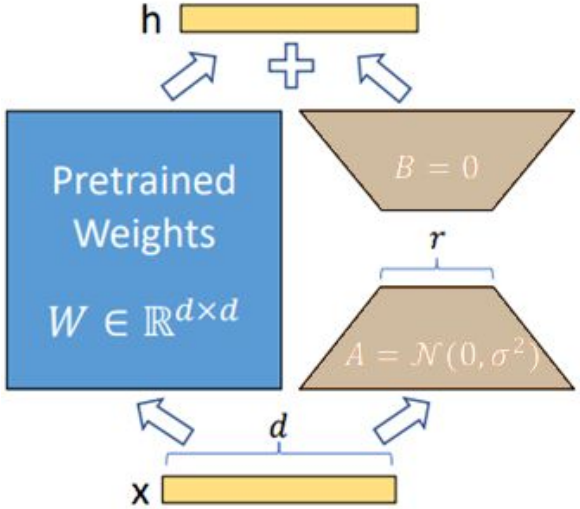
3. LoRA обучавање - експерименти

- Четири експеримента:
  - LoRA v1
  - LoRA v2
  - **LoRA v3**
    - DS_v2
    - train_batch_size: 8
    - unet_lr: 1e-4
    - text_encoder_lr: 5e-5
    - network_dim: 64
    - network_alpha: 64
  - LoRA v4



3. LoRA обучавање - експерименти

- Четири експеримента:
  - LoRA v1
  - LoRA v2
  - LoRA v3
  - **LoRA v4**
    - DS_v2
    - train_batch_size: 8
    - unet_lr: 1e-4
    - text_encoder_lr: 5e-5
    - network_dim: 64
    - network_alpha: 32



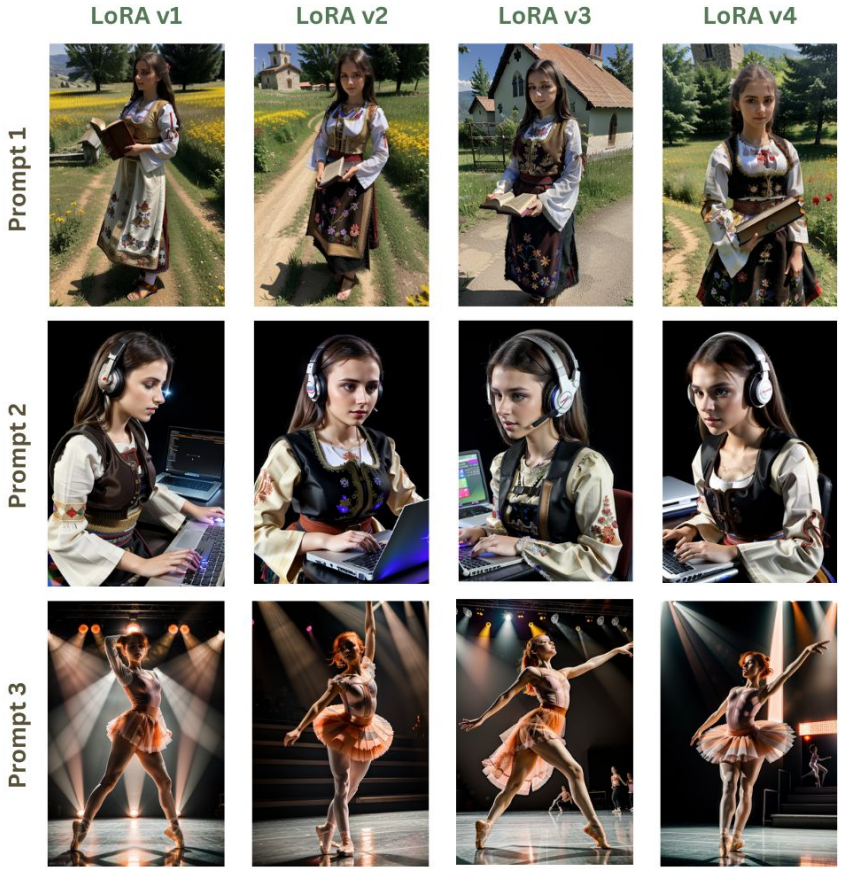
$$W' = W + \frac{\alpha}{r} \Delta W$$

4. Евалуација резултата

- Двојака евалуација:
  - Квалитативна - људским оком
  - Квантитативна - CLIP *score*

4. Евалуација резултата

- Двојака евалуација:
  - **Квалитативна - људским оком**
  - Квантитативна - CLIP *score*



4. Евалуација резултата

- Двојака евалуација:
  - Квалитативна - људским оком
  - **Квантитативна - CLIP *score***

$$CLIPscore(c, v) = w * \max(\cos(c, v), 0)$$

Prompt:

a girl wearing nosnjaoutfit, sunny day, city streets

LoRA v1

LoRA v2

LoRA v3

LoRA v4

no LoRA

4. Евалуација резултата

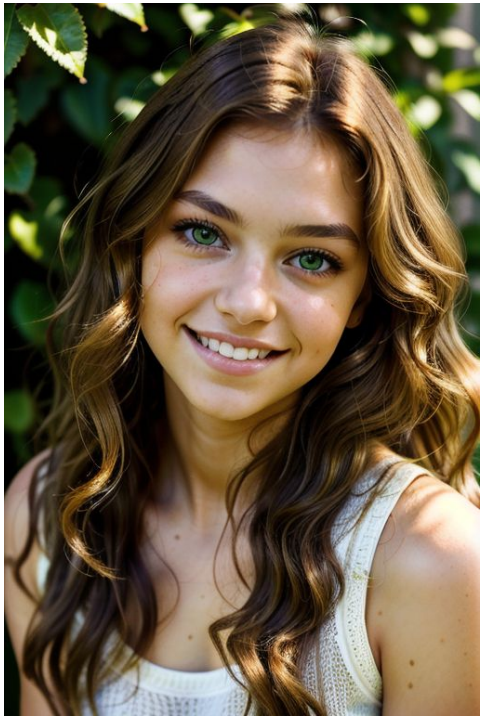
- Двојака евалуација:
  - Квалитативна - људским оком
  - **Квантитативна - CLIP *score***

$$CLIPscore(c, v) = w * \max(\cos(c, v), 0)$$

LoRA модел	*CLIP score*
LoRA v1	0.5786
LoRA v2	**0.6557**
LoRA v3	0.5894
LoRA v4	0.6101
no LoRA	0.5935

4. Евалуација резултата

- *trigger* реч



Дискусија и закључак

- 34 слике + 15 минута обучавања => успешна LoRA
- Величине фајлова до 72 MB => меморијска ефикасност
- Најбољи модел? Емпиријски одредити

Будућа унапређења

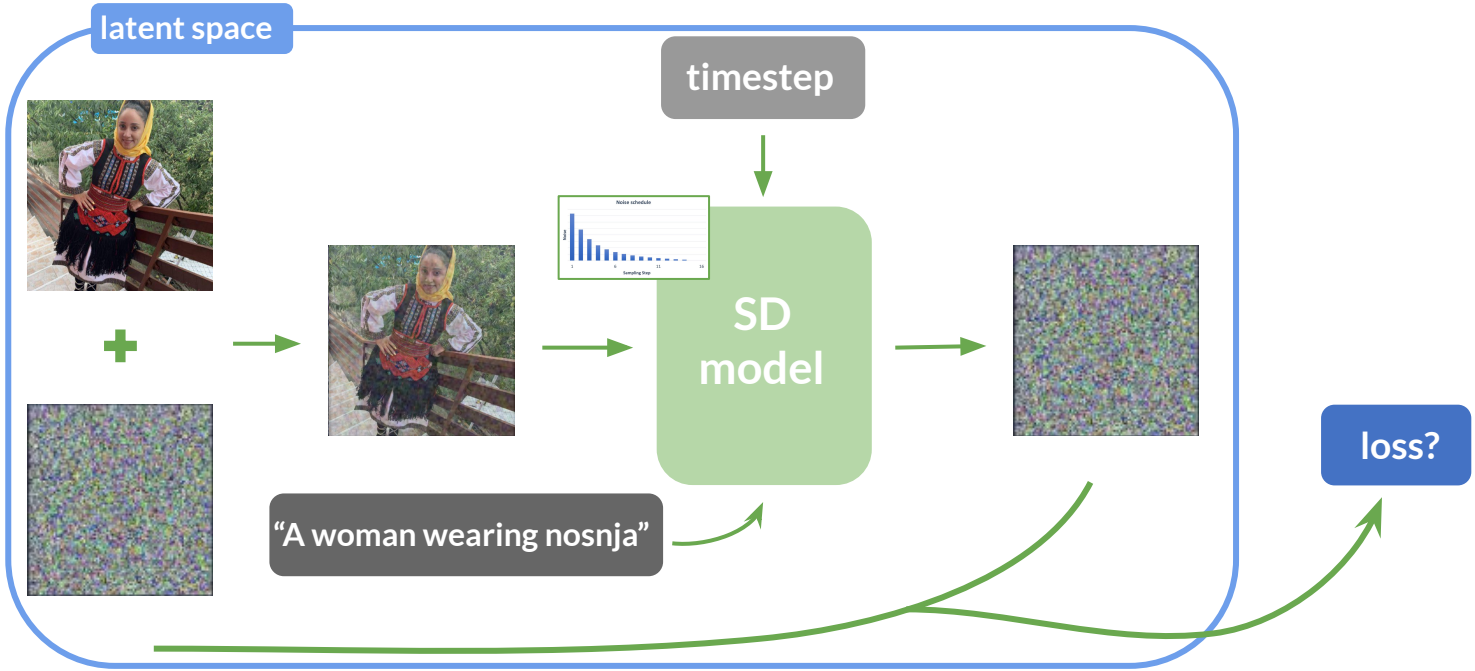
- Већи скуп података
- Регуларизација
- Хиперпараметри
- Квантитативна евалуација

Hvala na pažnji!

Pitanja? ✋

Stable Diffusion

- **Латентни** дифузни модел за генерисање слика.
- 48x мање бројева => брже и мање меморије заузима



Stable Diffusion

- **Латентни** дифузни модел за генерисање слика.
- 48x мање бројева => брже и мање меморије заузима