

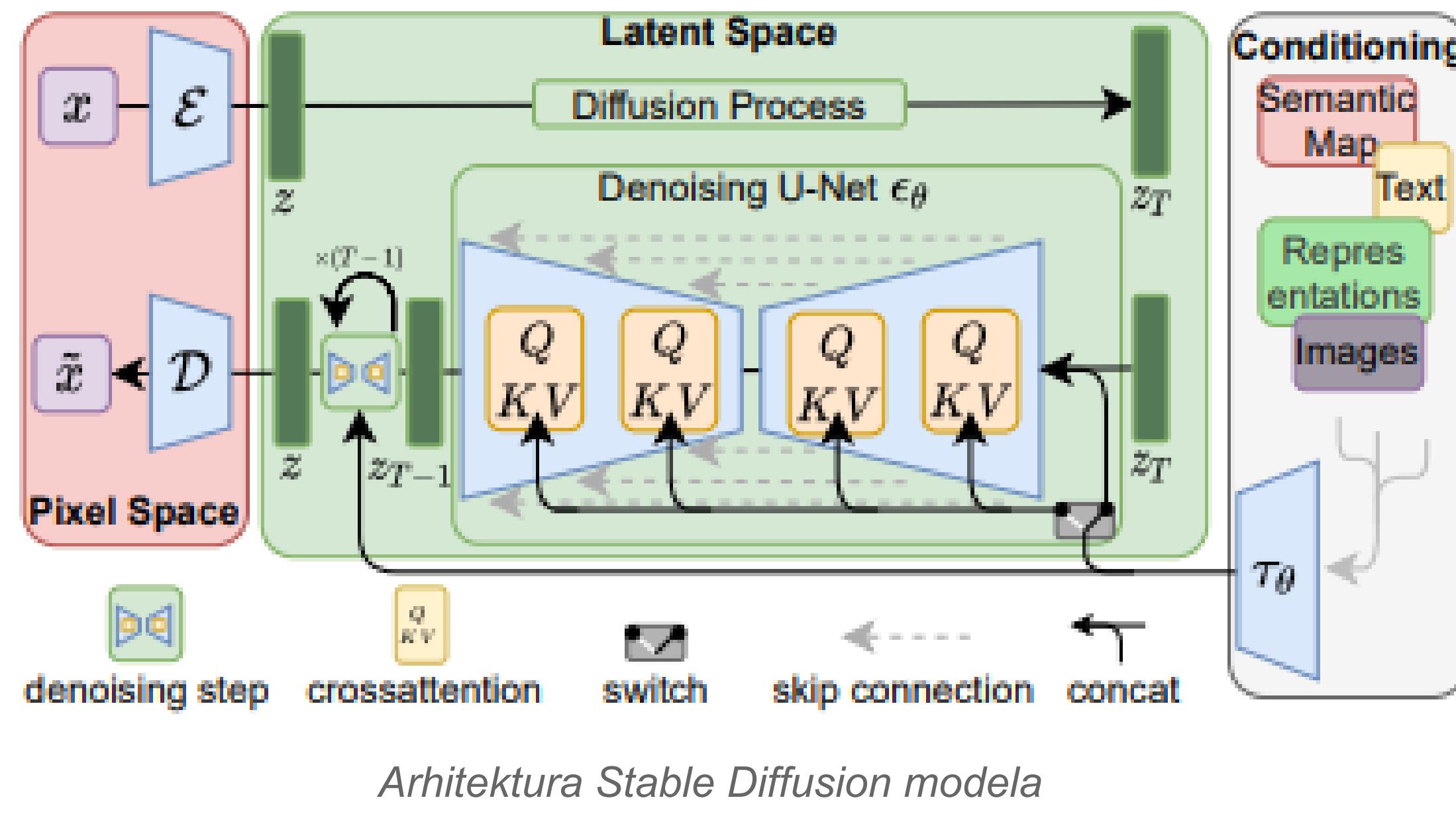
Fino podešavanje Stable Diffusion modela primenom LoRA metoda

LoRA za generisanje srpske narodne nošnje

Tina Mihajlović, Softversko inženjerstvo i informacione tehnologije

Uvod

Stable Diffusion je **latentni model** iz grupe **diffusion** modela, namenjih za generisanje slika na osnovu tekstualnih opisa. Razvijen je od strane StabilityAI-a, kao open source model sa performansama na **state-of-the-art** nivou.



Cilj projekta je upotreba jednog od metoda za fino podešavanje diffusion modela, pod nazivom **LoRA** (Low Rank Adaptation Matrices), kako bi se Stable Diffusion naučio da verodostojno generiše srpsku narodnu nošnju, što je koncept koji mu je pre treninga nepoznat.

Motivacija

LoRA metod podrazumeva zamrzavanje originalnih težina modela. Nove infomracije se uče kroz matrice koje se injektuju u *cross attention* slojeve, a čuvaju samo razliku originalnih i dotreniravanjem naučenih težina. Dakle, oslanjamо se na znanje koje model već ima i utičemo samo na deo njegovih težina, što čini ovaj metod veoma efikasnim sa strane RAM memorije i čuvanja izlaza treniranja - LoRA fajlovi su zauzimaju svega nekoliko stotina MB, a skraćeno je i vreme treniranja.

Metodologija

Projekat se može podeliti u dve semantičke celine:

- Prikupljanje skupa podataka
- Fino podešavanje modela

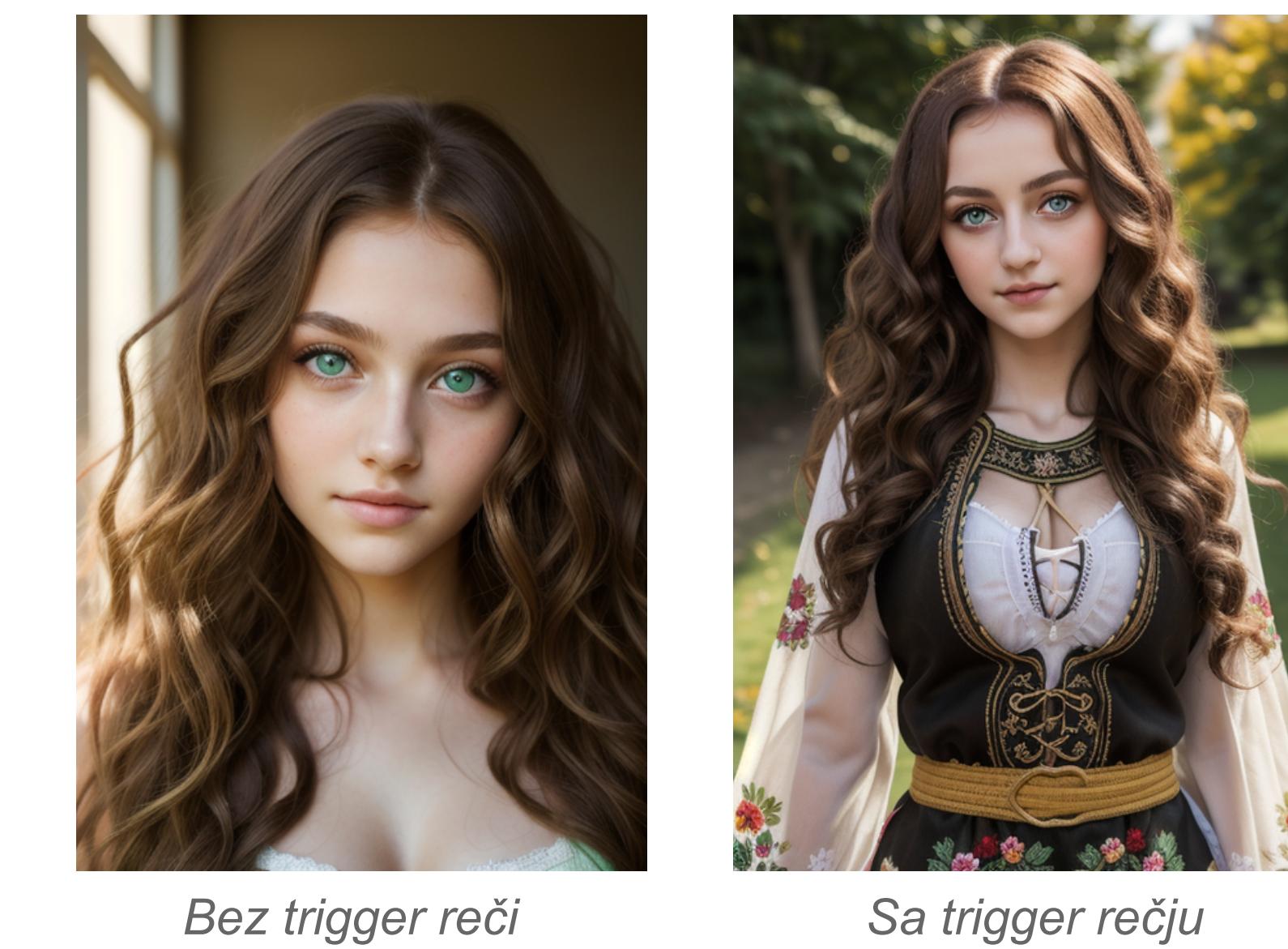
Prikupljanje podataka

Da bi se Stable Diffusion naučio konceptu kroz LoRA metod potrebno je svega nekoliko desetina slika - direktna posledica izloženog u segmentu "Motivacija". Međutim, ostaje da važi da što više kvalitetnih primera damo modelu, to će bolje naučiti koncept.

Kako nije pronađen odgovarajući postojeći skup podataka srpske narodne nošnje, ručno je prikupljen skup podataka za obučavanje.

Broj slika dovoljne rezolucije (bar 512x512px) nije bio znatan, pa je kroz augmentaciju podataka oformljen skup podataka od 34 slike.

Pored slika, modelu je potrebno proslediti i tekstualni opis svake slike, zašto je korišćen **BLIP model**, nakon čega su labele dopunjene trigger rečju "nosnjaoutfit" - **trigger reč** je reč koja bi trebalo da pobudi koncept koji učimo model, kada se iskoristi u *prompt*-u tokom *inference*-a.



Fino podešavanje modela & Rezultati

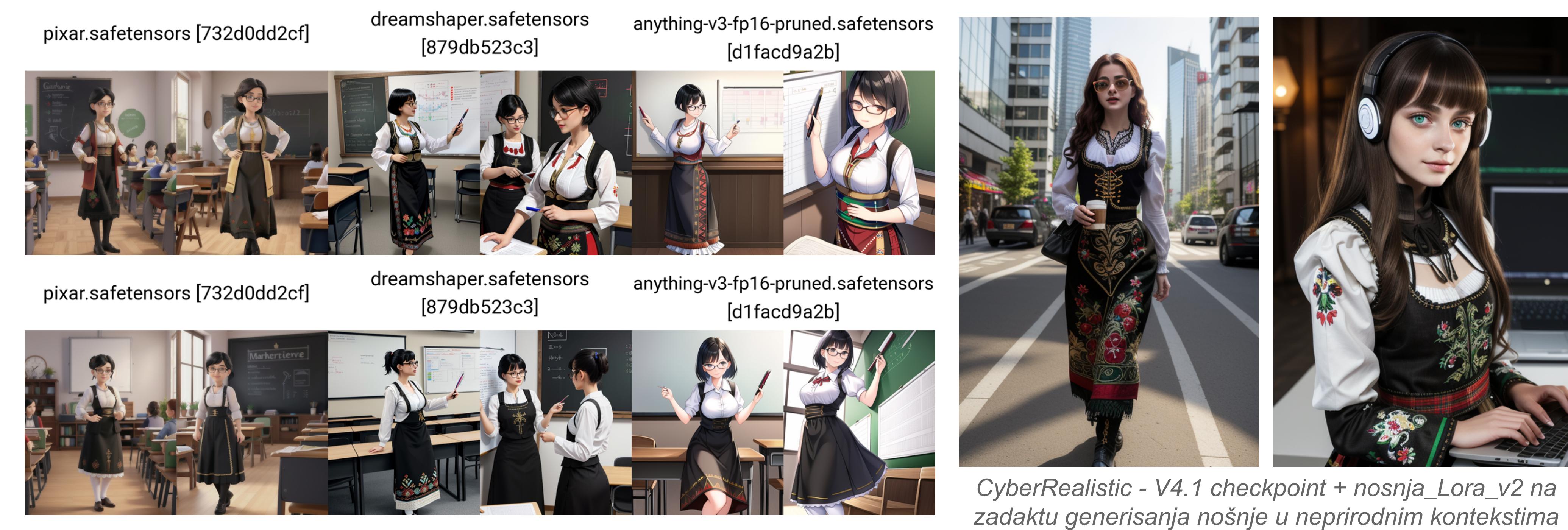
Sam proces treniranja izvršen je u *Google Colab* okruženju, sa *T4* grafičkom karticom. Treniranje je vršeno prilagođavanjem *Kohya* trener skripti, koje se oslanjaju na *HuggingFace* accelerator alat. Treniranje je izvršeno najpre na manjem skupu podataka od 20ak slika, i već tada su dobijeni zavidni rezultati:



CyberRealistic - V4.1 checkpoint + nosnja_Lora sa različitim prompt-ovima

Međutim, pri daljem testiranju se povremeno javlja problem koji je ukazivao na overfitting kroz lošu generalizaciju modela pri generisanju scenarija koji se jako razlikuju od trening skupa. Kroz povećanje veličine i raznovrsnosti skupa podataka, smanjenje *learning rate*-a i povećanje *batch size* parametra, razvijena je druga verzija (v2) LoRA-e, koja se bolje pokazala na testu generalizacije.

Uprkos tome, treba pomenuti da je prva verzija sa znacima overfittinga neretko generisala slike koje verodostojnije prikazuju srpsku nošnju od v2 verzije.



Poređenje nosnja_Lora (gore) i nosnja_Lora_v2 (dole) na identičnom zadatuču



CyberRealistic - V4.1 checkpoint + nosnja_Lora_v2 na zadaku generisanja nošnje u neprirodnim kontekstima (urban i futuristički)

Imajući u vidu specifičnost svakog testnog slučaja i uticaj hiperparametara poput seed-a, "bolju" od ove dve LoRA-e se predlaže odrediti u odnosu na konkretan zadatak, empirijski - obe omogućavaju da Stable Diffusion generiše i najfinije detalje srpske narodne nošnje, čak i kada se postavi u neprirodan kontekst i ili generiše u različitim stilovima.

Takođe, zaključena je i optimalna kombinacija parametara u opštem slučaju, kroz XYZ plot analize unutar AUTOMATIC1111 web aplikacije.

Zaključak

Uz primenu LoRA finog podešavanja, moguće je naučiti Stable Diffusion modele od milijardu parametara (v1) sa svega nekoliko desetina slika i 15ak minuta treniranja. Uz dobro pripremljen skup podataka, labele i eksperimentisanje sa različitim vrednostima hiperparametara, uspešno je postignut cilj generisanja srpske narodne nošnje, pri čemu se za dalje istraživanje ostavlja poboljšanje verodostojnosti i istraživanje povremenog problema zamućenih lica kao posledice promene težina.