



Střední průmyslová škola strojní
a elektrotechnická a Vyšší odborná škola,
Liberec 1, Masarykova 3

VYUŽITÍ STABLE DIFFUSION PRO VYTVÁŘENÍ ILUSTRAČNÍ GRAFIKY

Ročníková práce

Autor	Natálie Ryšavá
Obor	Technické lyceum
Vedoucí práce	Ing. Tomáš Kazda, DiS.
Školní rok	2022/2023

Anotace

Tato ročníková práce se zaměřuje na fascinující oblast generování obrázků pomocí umělé inteligence, která hraje v dnešním světě stále větší roli. Praktická část přináší komplexní návod, který čtenáře provede celým procesem generování obrázků ve Stable Diffusion web UI. Bez ohledu na to, zda jste umělec, student nebo manažer, tento tutoriál vám poskytne nezbytné znalosti pro vytvoření jakéhokoli obrazu, který si dokážete představit. Je navržen tak, aby byl srozumitelný i pro začátečníky bez předchozích zkušeností.

Summary

This work focuses on the fascinating field of image generation using artificial intelligence, which plays an increasingly key role in today's world. The practical part provides a comprehensive guide that guides the reader through the entire process of image generation in Stable Diffusion web UI. Whether you are an artist, student or manager, this tutorial will give you the necessary knowledge to create any image you can imagine. It is designed to be understandable even for beginners without previous experience.

Čestné prohlášení

Prohlašuji, že jsem předkládanou ročníkovou práci vypracovala sama a uvedla jsem veškerou použitou literaturu a bibliografické citace.

V Liberci dne 07.06.2023

.....

Natálie Ryšavá

Obsah

Úvod.....	1
1 Co je to neuronová síť?	2
1.1 Učení neuronové sítě	2
1.1.1 Učení s učitelem.....	3
1.1.2 Učení bez učitele.....	3
1.2 Využití neuronové sítě pro strojové učení	3
1.2.1 CLIP a difúzní modely.....	3
1.2.2 Příklady aplikací strojového učení	4
2 Prostředí Stable Diffusion web UI.....	6
2.1 Stable Diffusion checkpoint.....	7
2.2 Textové vstupy	7
2.2.1 Prompt.....	8
2.2.2 Negative prompt.....	15
2.2.3 Váha klíčového slova.....	18
2.3 Technologie generování obrázků.....	20
2.3.1 Text-to-image	20
2.3.2 Image-to-image.....	21
2.3.3 Extras	25
2.4 Sampling steps	27
2.5 Sampling method	28
2.5.1 Srovnání smplovacích metod	29
2.6 Rozměry	32
2.7 Batch count a batch size	33
3 Proces tvorby a optimalizace ilustrační grafiky.....	34
3.1 Člověk	35
3.2 Zvíře.....	37
3.3 Architektura	38
3.4 Krajina	40

Závěr.....	42
Seznam zkratek a odborných výrazů.....	43
Seznam obrázků.....	44
Použité zdroje	47
A. Seznam přiložených souborů	I

Úvod

Využila jsem příležitosti více se seznámit s novou průlomovou technologií umělé inteligence Stable Diffusion, která umožňuje generovat realistické obrázky.

Vývoj umělé inteligence se totiž v posledních letech velmi rychle rozvíjí a může být v budoucnu klíčovým faktorem v mnoha oblastech. Výzkumníci se snaží vylepšovat algoritmy, hardware a využívat nové technologie jako například kvantové výpočty a neuronové sítě. V budoucnu se očekává, že umělá inteligence přinese ještě rozsáhlejší vliv na lidský život a prokáže schopnost řešit náročnější problémy.

Stable Diffusion je otevřený text-to-image model založený na metodě učení CLIP, který dokáže generovat podrobné obrázky z libovolného textového vstupu. Byl vyvinut a vydán v roce 2022 společností StabilityAI ve spolupráci s EleutherAI a databází obrázků LAION. Model je primárně používán ke generování fotorealistických obrázků z latentního prostoru, upravování obrázků podle textu nebo doplňování chybějících detailů do snímků s nízkým rozlišením nebo nedostatečnou kvalitou. Byl natrénován na miliardách obrázků a dosahuje výsledků srovnatelných s DALL-E 2 nebo MidJourney.

Významný krok vpřed v oboru strojového učení představuje právě inovativní Stable Diffusion. Nabízí nové možnosti v oblasti umělé inteligence a může být využit mnoha způsoby, například při tvorbě filmových efektů, reklamních kampaní nebo v herním průmyslu.

Cílem je využít tuto práci jako učební pomůcku neboli návod pro studenty na to, jak používat model skrze webové rozhraní Stable Diffusion UI a implementovat vygenerované výstupy do svých projektů.

1 Co je to neuronová síť?

Neuronová síť (ANN neboli artificial neural network) je výpočetní model používaný v umělé inteligenci. Síťovou strukturou lze vyjádřit obecně nelineárním regresním modelem, který se používá k modelování vztahu mezi **vícerozměrnou vstupní proměnnou x** a **vícerozměrnou výstupní proměnnou y** . (1) (2)

Neuronové sítě můžeme definovat jako sadu algoritmů vytvořených podle fungování lidského mozku a nervového systému – pracují zkrátka na stejném principu jako trénovaný mozek dítěte. Skládají se z umělých (formálních) neuronů, což je jakási matematická funkce, která přijímá libovolný počet vstupů a poté je klasifikuje podle použitého algoritmu právě na jeden výstup. Každý neuron je tedy složen ze **vstupní vrstvy**, **více skrytých vrstev** a **výstupní vrstvy**. Ty jsou vzájemně propojeny synaptickými vazbami vytvářenými a upravenými pomocí strojového učení, což umožňuje neuronovým sítím adaptovat se na data a zlepšovat svou schopnost provádět úlohy, jako je **klasifikace** či **predikce**. Mezi sebou si vrstvy navzájem předávají signály a ty pak transformují pomocí aktivačních přenosových funkcí. (2) (3) (4)

1.1 Učení neuronové sítě

Pro učení neuronové sítě se musí nejprve vytvořit struktura sítě, která slouží jako základ pro její následné trénování. Nicméně, sama o sobě tato struktura neumí v podstatě nic, dokud není trénována na konkrétní úlohu. (5)

Trénování neuronové sítě spočívá v přijímání obrovského množství dat a postupném zlepšování její schopnosti interpretovat a chápat tato data. Přičemž během tohoto procesu se síti podávají různé podněty, které ji vedou právě k vhodnému porozumění předložených informací. V průběhu trénování si síť vytváří model, což je v podstatě abstraktní reprezentace vnitřních parametrů a vah, který se poté používá ke klasifikaci dat nových. Po dokončení trénování se síť stává aktivní, kdy může být využita pro řešení zadaných úloh. (3) (6)

Existují různé výukové metody neuronových sítí, které se liší v závislosti na typu úlohy, datového souboru a požadovaném výstupu. Klasickými jsou dvě paradigmaty typická pro strojové učení: **učení s učitelem** a **učení bez učitele**. (7) (8)

1.1.1 Učení s učitelem

Učení s učitelem, známé také jako „supervised learning“, představuje třídu postupů trénování dat obsahujících vstupní objekty a odpovídající výstupní ohodnocení. To umožňuje strojům učit se z dat, která jsou již kategorizována nebo anotována lidmi pomocí klíčových slov, a tyto vzory pak slouží jako referenční bod. Proces probíhá tak, že stroj obdrží trénovací data, porovná je a snaží se najít funkci, která nejlépe popisuje jejich závislost mezi vstupem a výstupem. Na základě tohoto tréninku by měl s velkou přesností umět definovat data nová. (9) (10) (11)

Například při trénování obrazové klasifikace by se síti mohly prezentovat různé obrázky spolu s příslušnými texty, které specifikují, co se na obrázku nachází. Síť pak adaptivně upravuje své vnitřní parametry tak, aby příště dokázala vstupní data správně vyhodnotit. (12)

1.1.2 Učení bez učitele

Učení bez učitele neboli „unsupervised learning“ se od učení s učitelem liší tím, že trénovací data neobsahují výstupní ohodnocení. To znamená, že funkce musí najít vztahy v datech sama metodou pokus-omyl. Využívá shlukování dat, aby objevila podobnosti mezi různými objekty, přiřadila jim vlastní označení a zařadila je do skupin se srovnatelnými vlastnostmi. (8) (11)

1.2 Využití neuronové sítě pro strojové učení

V dnešní době existuje mnoho známých nástrojů strojového učení, které přinášejí inovativní a užitečné řešení pro různé oblasti. Například to je rozpoznávání obrazů a řeči, klasifikace a segmentace dat, předpovídání vývoje časových řad, analýzy psaného textu nebo filtrování spamu. V lékařství se používá k diagnostice onemocnění a v řízení pro podporu rozhodování. (8) (11)

1.2.1 CLIP a difúzní modely

CLIP (Contrastive Language-Image Pre-training) představuje model neuronové sítě vyvinutý společností OpenAI, který kombinuje obrazovou a jazykovou reprezentaci. Je vytrénován na více než 400 milionech dvojic textu a obrázku, a to například pomocí **difúzních modelů**. Tradičně bylo nutné pro kategorizaci shromáždit velké množství

označených obrázků a na nich natrénovat klasifikační model. CLIP je však zero-shot model, což znamená, že dokáže identifikovat nespočet věcí, které nikdy předtím neviděl. (13) (14) (15)

Difúzní modely jsou pokročilé strojově učené algoritmy pro generování dat, které využívají **princip difúze** k šíření informací a provádění změn v datových strukturách. Zpracovávají obrazy (případně i videa) s cílem odhadnout a předpovědět neznámé hodnoty na základě těch známých.

Během trénování se difúzní modely učí odstraňovat šum, který je zaveden do obrázků. Tento proces obnovy šumu je pak využíván při generování obrazů zcela nových, kde model vytváří realistické vizuální obsahy z původních nejasností. Díky principu difúze tyto modely dokáží efektivně šířit a propagovat informace skrze datové struktury, což přináší vylepšené výsledky při práci s obrazovými a video daty. (16) (17)

1.2.2 Příklady aplikací strojového učení

Mezi nejpopulárnější aplikace strojového učení patří **generátory obrázků na základě textových popisů** nebo **chatovací nástroje**. Mohou být buď zdarma a volně dostupné pro veřejnost, nebo komerční a vyžadovat placenou licenci. Výběr konkrétní aplikace závisí na individuálních potřebách a požadavcích uživatele. Nejznámějšími generátory obrázků jsou **DALL-E 2**, **Imagen**, **Midjourney** a **Stable Diffusion**, které sice slouží ke stejnému účelu, avšak přinášejí odlišné výsledky. Umí například tvorbu jakéhokoli stylu, třeba i podle konkrétního umělce. K významným chatovacím nástrojům naopak lze přiřadit chatbot **ChatGPT**. (18) (19) (20)

1.2.2.1 DALL-E 2

Výsledky z DALL-E 2 nedosahují takové úrovně jako jiné AI generátory. Co se týče uměleckého ztvárnění, není to jeho nejvýraznější vlastnost, chybí mu flexibilita a pokročilé funkce. Na druhou stranu exceluje v oblasti fotorealismu. Byl vyvinut společností **OpenAI**, na jejichž webu je dostupný veřejnosti s omezeným množstvím bezplatných obrázků. Od dubna 2022, kdy byl odhalen, poskytuje ještě realističtější obrázky s 4× vyšším rozlišením než jeho původní verze DALL-E 1. (21) (22)

1.2.2.2 Imagen

Imagen reprezentuje jakousi verzi DALL-E 2 stvořenou **Googlem**, ale zatím není dostupná pro veřejnost – přístup mají pouze zaregistrovaní beta testéři. Je totiž trénována na obrovských souborech dat stažených z webu, které nejsou upravovány, což přináší řadu problémů. Chlubí se nevídaným stupněm fotorealismu a hlubokým porozuměním. Dokáže také automaticky rozpoznat obsah obrázků či videí a přiřadit jim metadata. Staví na síle velkých transformačních jazykových modelů. (19) (23)

1.2.2.3 Midjourney

Midjourney je inovativní a **nezávislý výzkumný projekt**, jehož cílem je rozšířit lidskou představivost a otevřít nové možnosti ve vytváření vizuálního umění. Vyniká v tvorbě obrazů, které jsou velmi podobné dílům skutečných umělců. Momentálně je platforma stále ve fázi výzkumu a vývoje, proto je přístupnost omezená. Uživatelé mohou vyzkoušet tuto technologii pouze prostřednictvím bota po registraci na oficiálním Discord serveru. Nabízí zkušební verzi zdarma. Pro více funkcí a vyšší množství obrázků je nutné si připlatit, díky čemuž uživatel získá také právo využívat výstupy komerčně. Nedávný případ, kdy někdo v soutěži využil vygenerovaný obrázek z Midjourney a vyhrál, vyvolal otázky ohledně autorských práv a férovosti. (24) (25)

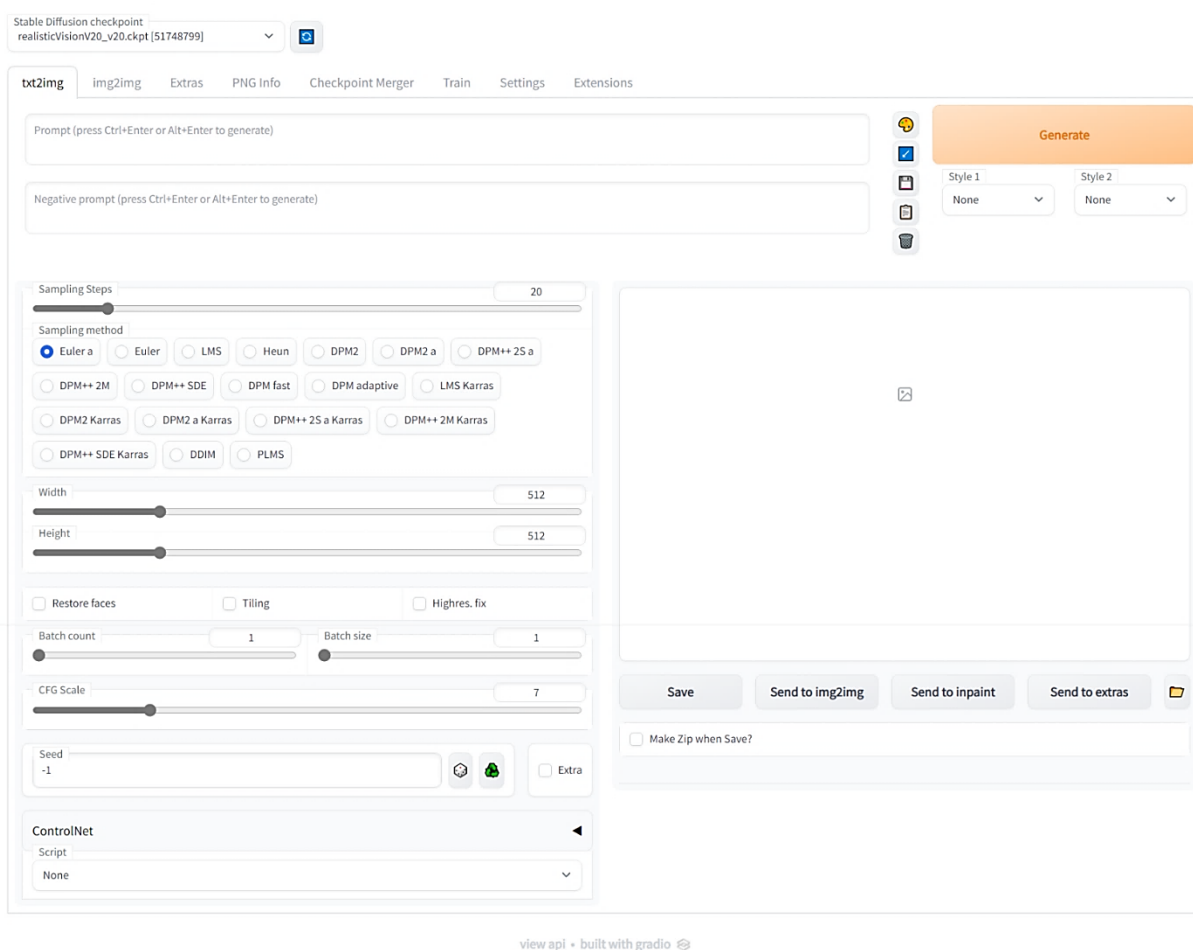
1.2.2.4 Stable Diffusion

Stable Diffusion, vydán společností **StabilityAI** v srpnu 2022, je jediný open-source model z předchozích zmíněných. Představuje však největší výzvu v dosažení požadovaného výsledku, kterému mají pomoci tzv. negativní prompty. Trénink byl realizován pomocí CLIPu ve spolupráci s řadou akademických výzkumníků a neziskových organizací. Hlavním účelem je generovat detailní obrázky na základě textových popisů, ale lze ho využít i pro další úkoly. Uživatelé mají možnost spustit ho lokálně na všech počítačích vybavených GPU s minimálně 8 GB VRAM přes Stable Diffusion web UI. To poskytuje větší kontrolu, lepší přizpůsobení, a dokonce umožňuje trénování modelu vlastního. (26) (24) (27)

Díky veřejnému zdrojovému kódu vzniklo také několik nástrojů umožňujících přístup k technologii Stable Diffusion přímo z webového prohlížeče. Například platforma Lexica využívá svůj vlastní model nazvaný „Aperture“, který se vyznačuje vysokou kvalitou a mírou důvěryhodnosti. (19)

2 Prostředí Stable Diffusion web UI

Stable Diffusion web UI je prostředí, které slouží k vizualizaci a správě rozsáhlých sítí a datových toků. Toto uživatelské rozhraní představuje klíčový nástroj pro interakci s modelem Stable Diffusion a umožňuje primárně efektivně generovat obrázky, analyzovat je a provádět s nimi různé jiné operace. Hlavními prvky jsou navigační lišta, menu, pracovní plocha a panel s nástroji. To usnadňuje uživatelům orientaci a manipulaci při generování.



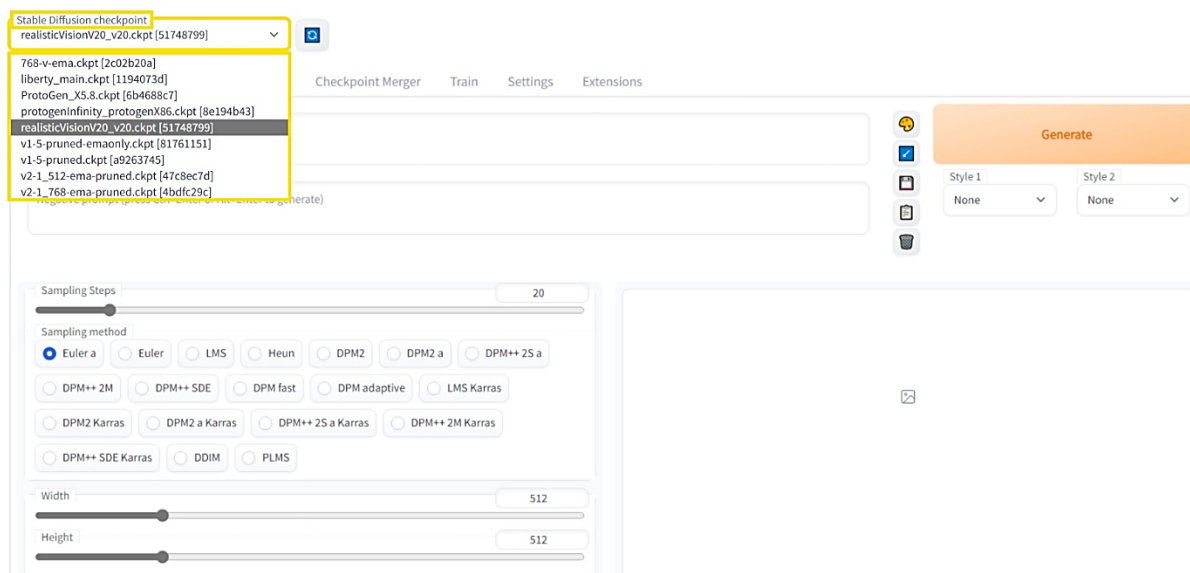
Obrázek 1 Prostředí Stable Diffusion web UI

Pro svůj provoz využívá tato platforma moderní technologie a frameworky. Je optimalizována pro podporu různých webových prohlížečů a zařízení, což poskytuje uživatelům flexibilitu a dostupnost na různých platformách. Doporučuje spouštět s VRAM o kapacitě 10 GB nebo vyšší, nicméně uživatelé s nižší kapacitou mohou na úkor výkonu modelu zvolit jiné načítání vah a Stable Diffusion také zapnou.

Společnost StabilityAI, jež model vyvinula, si nenárokují žádná práva na generované obrázky a bezplatně uděluje uživatelům oprávnění k jejich použití. Avšak pouze za předpokladu, že obsah není nelegální ani škodlivý. Svoboda, kterou poskytuje občas vyvolává kontroverze ohledně etiky vlastnictví, neboť model je trénován na základě autorských obrázků bez souhlasu jejich majitele, jak již bylo zmíněno v kapitole 1.2.2.4. (27) (28)

2.1 Stable Diffusion checkpoint

Vlevo nahoře se nachází karta, která poskytuje přehled všech dostupných „kontrolních bodů“ modelu. StabilityAI v srpnu 2022 vydala první veřejný checkpoint Stable Diffusion v1.4. V následujících měsících pokračovala ve vydávání verzí 1.5, 2.0 a 2.1.



Obrázek 2 Přehled Stable Diffusion checkpointů

Tyto checkpointy určují, z jakých obrázků bude prováděn výběr pro generování. Každý jednotlivý model je totiž natrénován na sadě obrázků jiného typu. Proto se některý hodí spíše pro fotorealistické snímky, zatímco jiný je vhodnější pro malbu nebo digitální díla. (29) (30)

2.2 Textové vstupy

Do Stable Diffusion lze zapsat dva typy textových vstupů: **prompt** a **negative prompt**. Při jejich samotnému psaní se musí dbát na několik zásad.

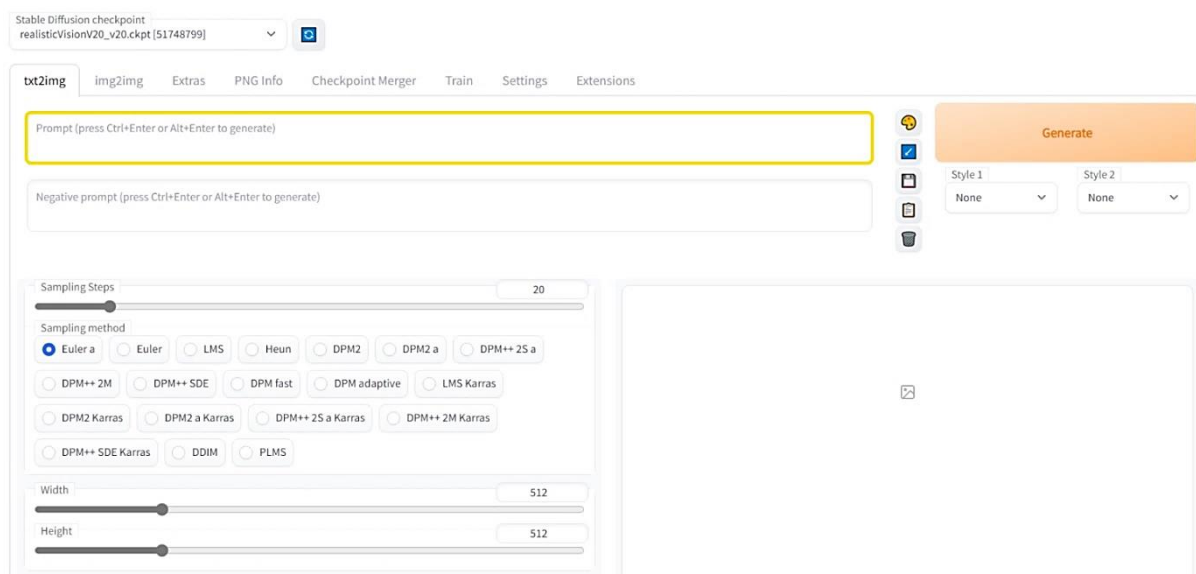
Prvně je nutné stanovit si přesnou specifikaci tématu nebo motivu, aby byly vygenerované obrázky co nejrelevantnější. Poskytnutí kontextu umožňuje generátoru porozumět účelu promptu a vytvořit smysluplné umělecké dílo. Nezbytností je psát texty **v anglickém jazyce**, neboť na něm byl model natrénován. Zároveň je zapotřebí vynechat všechna výplňová slova a vyvarovat se nepředmětným nebo neobjektivním dlouhým frázím, které by mohly omezit tvůrčí svobodu umělce a tím i rozsah promptu. Naopak se doporučuje používat neutrální výraziva, která povzbuzují objevování a experimentování. Místo dlouhého spojení slov „*oil painting of an apple on table by Vincent van Gogh*“ by měla být použita jednoduchá slova s přívlastky jako „*oil painting, apple on table, by Vincent van Gogh*“.

Dalším důležitým prvkem je používání otevřených otázek, které poskytují prostor Stable Diffusion k rozvoji kreativity a příchodu s originálními řešeními. Lze tak vyjádřit hlubší myšlenky a emoce umělce.

Nakonec je nutné otestovat své prompty před jejich finálním použitím. Testování totiž zaručuje jejich jasnost, stručnost a účinnost při generování požadovaných odpovědí. Tímto způsobem se dosáhne velmi precizního výsledku. (31) (26) (32) (33)

2.2.1 Prompt

Prompt je textový vstup, který poskytuje modelu pokyny či instrukce ohledně **požadovaného** obrázku. Slouží k zahájení generování, definuje obsah a styl. Prompty mohou být specifické nebo otevřené v závislosti na smýšleném výstupu. Je však nutné mít na paměti, že i jejich délka může mít vliv. Delší prompty zajišťují modelu více kontextu a podrobností, zatímco ty kratší mohou vést ke kreativnějším a rozmanitějším reakcím. Někdy ale udělá úžasnou práci a vygeneruje přesně to, co uživatel chce, s nejasným promptem. Jindy dodá suboptimální výstupy.



Obrázek 3 Příkazový řádek pro prompt

Správný prompt by měl obsahovat minimálně tyto kategorie klíčových slov: **předmět, médium, styl, umělec, webové stránky, rozlišení, barva, osvětlení a další podrobnosti**. Tyto informace přispějí k lepšímu porozumění generátoru a vytvoření požadovaného obrázku. (31) (19) (32) (34) (35)

2.2.1.1 Předmět

Předmět je hlavní obsah, který chceme na obrázku vidět. Častou chybou činí jejich nedostatečný popis.

Pro lepší představu lze zvolit postavu mocné královny. Před zahájením tvorby by mělo být zodpovězeno uživatelem několik základních otázek, kterými se bude řídit při psaní promptu.

- Jak vypadá?
- Co má na sobě?
- V jakém prostředí se nachází?
- Jaké má stáří?
- Jakou činnost vykonává?
- V jaké je poloze?

Výsledným promptem tedy jsou odpovědi na tyto otázky – „*queen with blue dress, crown, sitting on red throne, young, beautiful, powerfull*“.



Obrázek 4 Obrázek mocné královny vytvořený ve Stable Diffusion

Běžným trikem pro generování postav je používání jmen celebrit. Mají silný účinek a jsou vynikajícím způsobem, jak ovládat vzhled. Mohou změnit nejen obličej, ale třeba i pózu. K promptu na vygenerování obrázku královny bylo přidáno jméno známé herečky Emmy Watson – „queen with blue dress, crown, sitting on red throne, young, beautiful, powerfull, Emma Watson“.



Obrázek 5 Obrázek Emmy Watson jako mocné královny vytvořený ve Stable Diffusion

2.2.1.2 Médium

V případě umění médium znamená materiál neboli prostředek vyjádření, jakým jsou vytvářena umělecká díla. Může být viděno několika různými způsoby a zcela závisí na kontextu. Média zprostředkovávají umělcův nápad, poselství či podnět. Příkladem je ilustrace, olejomalba, 3D renderování, kreslená pohádka, japonské anime a realistická fotografie, případně i typ fotoaparátu. Použití je silně efektivní, protože samo o sobě jedno klíčové slovo dokáže dramaticky změnit celý dojem z obrázku.

Přidáním jednotlivých výrazů „oil painting“, „anime“ a „RAW photo, dslr, Fujifilm XT3“ do existujícího vzorového promptu s královnou vznikla tři zcela odlišná díla



Obrázek 6 Olejomalba mocné královny vytvořená ve Stable Diffusion



Obrázek 7 Anime obrázek mocné královny vytvořený ve Stable Diffusion



Obrázek 8 Realistická fotografie mocné královny vytvořená ve Stable Diffusion

2.2.1.3 Styl

Styly jsou důležitou součástí tvorby a porozumění uměleckým dílům. Jejich zmínka v promptu umožňuje široké spektrum tvůrčích perspektiv a otevírá prostor pro kontrasty mezi jednotlivými směry. Představují velkou škálu výrazových forem, které reflektují umělecké hnutí, dobu a individuální tvůrčí přístupy umělců. Uvádění těchto stylů má mnoho výhod. Tím, že vytyčují konkrétní směr, umožňují generovat cílenější výstupy.

Každý jednotlivý styl ukazuje specifické estetické a konceptuální prvky, jako je snovost a iracionalita v surrealismu nebo typická geometrická abstrakce v kubistických dílech. Nicméně styly nemusí být čistě umělecké směry, může se například jednat o „hyperrealistic“ nebo zobrazení celé postavy na obrázku „full body“. Často užívaným je také komiksově laděný „pop-art“ nebo „fantasy“ užívající magii a jiné nadpřirozené prvky.

K ukázkovému promptu lze přidat například „*Disney, fantasy*“. Pro lepší výsledek je doplněno také médium „*cartoon*“, které umožňuje do děl vnést skutečnou pohádkovou tematiku.



Obrázek 9 Fantazijní obrázek mocné královny v tématice kreslených Disney pohádek vytvořený ve Stable Diffusion

2.2.1.4 Umělec

Jména umělců mají silný vliv na tvorbu a mohou sloužit jako klíčové inspirační prvky v rámci generování obrázků ve Stable Diffusion. Poskytují možnost vybrat si přesný styl díky konkrétnímu umělci. Často se také kombinují různá jména umělců, aby se propojily jejich unikátní styly.

V současnosti je jedním z nejvíce užívaných umělců **Greg Rutkowski**, talentovaný digitální malíř s nezaměnitelným stylem. Jeho obrazy často zachycují magické scenérie, kde se setkává skutečnost s imaginací. Případně lze přidat také dvě znamenitá jména, a to Stanleyho Artgerma Laua, mistra komiksového kreslení superhrdinů, a Alphonse Muchu, vynikajícího portrétistu 19. století. U realistických fotografií jsou hojně používáni talentovaní tvůrci Annie Leibovitz, Steve McCurry nebo Richard Avedon. Jejich příspěvky jsou neocenitelné a zanechávají trvalou stopu ve světě fotografie.

Tentokrát bylo při generování využito „*by Leonardo da Vinci*“ ve spojitosti s „*oil painting*“. Leonardo da Vinci je autorem nejslavnějšího obrazu všech dob, portrétu zvaného Mona Lisa, který také generovaný obrázek připomíná.



Obrázek 10 Olejomalba mocné královny inspirovaná díly malíře Leonarda da Vinciho vytvořená ve Stable Diffusion

2.2.1.5 Webové stránky

Stable Diffusion využívá různé zdroje na internetu pro získání inspirace při tvorbě obrázků. Existuje totiž řada specializovaných grafických webových stránek, jako je Artstation a Deviant Art, které slouží jako sběrnice pro množství obrázků různých žánrů. Tyto platformy poskytují ideální zdroj nápadů pro prompty, umožňují nasměrovat tvorbu obrázků do těchto specifických stylů. Mezi další stránky patří také Behance, Flickr, Unsplash, Instagram nebo Pinterest sdílející nápady ve formě "pinů" a Dribbble zaměřující se především na design.

Výsledkem je zlepšená kompozice, barevná paleta nebo jiné jemné prvky, které se promítají do celkové kvality a estetiky snímku. I když tyto změny nemusí být okamžitě viditelné na jednotlivých obrázcích, celkový efekt a zdokonalení může být patrné při porovnání původních a vylepšených verzí.

2.2.1.6 Rozlišení

Rozlišení ukazuje míru, jak detailní a ostrý je obraz. Prompty používají klíčová slova, která zdůrazňují požadovaný efekt. Například fráze „*highly detailed*“, „*high resolution*“, „*ultra quality*“, „*crystal clear*“ či „*sharp focus*“ instruují algoritmus tak, aby se zaměřil na získání nejvyššího možného stupně rozlišení a ostrých kontur výsledného snímku.

Mohou se použít i parametry fotoaparátu nebo konkrétní druhy objektivů s jejich clonovým číslem. Například „50 mm f2.8“ pro klasický portrétní vzhled s malou hloubkou ostrosti, „85 mm f1.8“ pro lichotivou kompresi obličejových rysů nebo „200 mm f2.8“ pro fotografování zvěře v dále.

Kombinací těchto promptů s dalšími a následnými úpravami je podtrženo maximální množství jemných detailů a perfektního zaostření. To přidává generátoru další variabilitu.

2.2.1.7 Barva

Vizuální přitažlivost obrázku je možné vylepšit začleněním slov označujících různé barvy. Zadáním takových promptů lze do generovaného výstupu zavést specifické tóny nebo zvýraznit jednotlivé objekty odlišnými odstíny tak, aby vyhovoval různým náladám a tématům. Nádech zářivosti naznačuje „*iridescent gold*“, intenzivní odstín červené „*vibrant red*“, pocit mystiky vytváří „*radiant purple*“ a v neposlední řadě například „*pearl white*“ nebo „*pastel colors*“ dodává obrazu elegantní a nadčasový nádech.

Barva bývá popsána často již při představení hlavního předmětu. Názornou ukázkou slouží „*queen in **blue** dress*“ a „*sitting on **red** throne*“ v textovém vstupu u generování mocné královny.

2.2.1.8 Osvětlení

Nejenom fotografové, ale všichni tvůrci vyzdvihují, že osvětlení je základním faktorem při vytváření úspěšných snímků a děl. Taková klíčová slova mají obrovský vliv na celkový vzhled, náladu a atmosféru, přičemž přináší obrázkům zcela nový rozměr a hloubku.

Kromě tradičních termínů, jako je „*cinematic lightning*“ nebo „*studio lightning*“, existuje celá řada dalších možností, které mohou sloužit jako inspirace. Například prompty jako „*dark*“ nebo „*mystical lighting*“ by mohly evokovat obrazy tajemných stínů a dramatických kontrastů, které předají hluboký emocionální dotek. Mezi jemnější a subtilnější osvětlení se řadí „*soft glow*“ nebo „*gentle lighting*“, která vyvolají pocit nostalgie nebo romantiky. V takových případech lze zkusit i „*soft lightning*“, což je velmi hojně užíváno u fotografií, především těch portrétních. Vystupuje zjemnělá paleta barev, zvýrazní se důležité prvky a vytvoří se snová atmosféra. Dalším příkladem je použití

promptu „*dramatic lighting*“, což může vést k výrazným a kontrastním obrazům s jasně osvětlenými objekty. To dodává generovaným výstupům dynamiku a intenzitu.

Velký rozdíl lze pozorovat při porovnání dvou fotografií s prompty identifikujícími odlišné světelné podmínky – nejprve „*soft lighting*“ a poté „*dramatic lighting, dark*“. Dramatizující pozadí u druhého obrázku dokonce tvoří modrý blesk.



Obrázek 11 Realistická fotografie mocné královny s měkkým osvětlením vytvořená ve Stable Diffusion



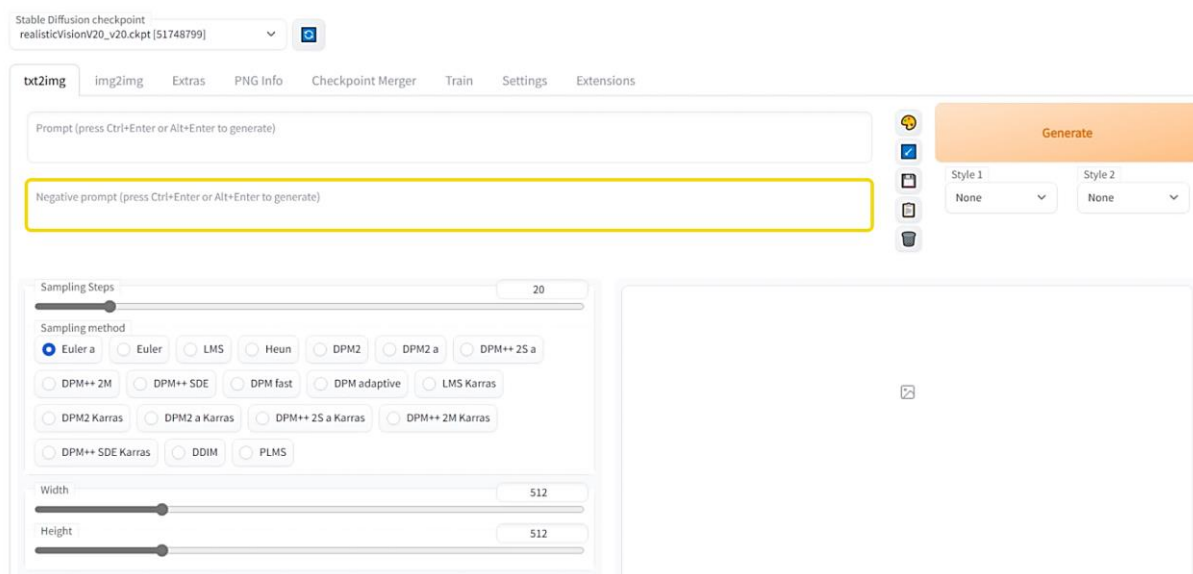
Obrázek 12 Realistická fotografie mocné královny s tmavým dramatickým osvětlením vytvořená ve Stable Diffusion

2.2.1.9 Další podrobnosti

Ve Stable Diffusion jde kromě základních informací přidávat do promptu dodatečné prvky a podrobnosti, které přispívají k vytvoření ještě výraznějších a poutavějších obrázků. Třeba začlenění termínů jako „*sci-fi*“, „*stunningly beautiful*“ a „*dystopian*“ přenáší do děl futuristický, nádherný nebo temný pocit. Je vhodné detailněji popsat prostředí, postavy, objekty, kompozici, perspektivu, a dokonce i jiné speciální efekty. Uživatelé se mohou zaměřit na cokoli, co jim přijde na mysl. Například lze přidat termín „*low angle*“, což vytváří dramatický efekt z pohledu zdola, „*elegant*“ pro vyjádření vkusu a sofistikovanosti nebo „*exotic*“ pro zobrazení neobvyklých a cizokrajných prvků.

2.2.2 Negative prompt

Negativní prompt je textový vstup, který obsahuje informace o tom, co by na výsledném díle **nemělo být přítomno**. V podstatě určuje, co uživatel chce potlačit či z obrázku odstranit. Slouží k zahájení generování stejně jako „pozitivní“ prompt, ale se zaměřením na formulaci atributů nebo stylů, které mají být minimalizovány.



Obrázek 13 Příkazový řádek pro negativní prompt

V kontextu vizuálního zpracování je klíčovým prvkem procesu difúze, protože zabraňuje vzniku mutací, deformací, malformací, duplikátů a jiných nežádoucích prvků. Tím přispívá také ke zlepšení kompozice a estetického dojmu výsledného díla. Namísto výrazů „sharp“ nebo „focused“ v promptu, může být v negativním promptu „blurry“. Obraz se stává ostřejší. Nezbytnost užití spočívá ve schopnosti vytvářet obrazy s vyšší kvalitou a vizuálním účinkem, což Stable Diffusion charakterizuje.

Ve skutečnosti lze některé obrázky generovat pouze s pomocí negativních promptů, k čemu poslouží názorné ukázky. Třeba v případě, kdy byla generována mocná královna, měla téměř vždy nějaký náhrdelník. Když by chtěl uživatel jeho tvorbu eliminovat, možná by bez rozmyslu ke stávajícímu promptu poskytl pouze „without necklace“, což však Stable Diffusion ve většině případů neuposlechne. To proto, že chápe negativní prompty spíše než lidský jazyk a jeho zápor. Na vině je tedy pravděpodobně selhání při asociaci „without“ a „necklace“.

- **Prompt:** queen with blue dress, crown, sitting on red throne, young, beautiful, powerfull, dark background, dslr, RAW photo, ultra quality, 8K UHD, high detailed, soft lighting, (skin pores, high detailed skin, sharp focus:1.2), film grain, Fujifilm XT3 + without necklace



Obrázek 14 Realistická fotografie mocné královny bez "without necklace" v promptu vytvořená ve Stable Diffusion



Obrázek 15 Realistická fotografie mocné královny s "without necklace" v promptu vytvořená ve Stable Diffusion



Obrázek 16 Realistická fotografie mocné královny s "necklace" v negativním promptu vytvořená ve Stable Diffusion

Zjevně lze pozorovat, že všechny fotografie sdílejí určitý charakteristický rys. Projevuje se v jejich nedokonalosti, kde postavy mají deformované končetiny a obličeje, přičemž se také vyskytují opakující se prvky, jako jsou koruny nebo dokonce postava královny, která se objevuje na obraze. Tento jev nastal v důsledku chybějících negativních promptů až na jednu výjimku u posledního obrázku, kde je patrná absence náhrdelníku. Tyto snímky tedy nejsou vhodné k použití.

Aby fotografie královny plnila svůj účel, je vhodné použít více negativních promptů zvýrazňujících již zmíněné deformace a duplikace. Vzhledem k přítomnosti malířských a polorealistických znaků je nezbytné omezit i to. V následující kapitole 2.2.3 je vysvětleno, proč jsou v textových vstupech obsaženy závorky s čísly. Využití negativních promptů u lidských postav pro dosažení optimálních výsledků se podrobněji rozebírá v kapitole 3.1. Proces vyžaduje několik zkušebních generování s cílem nalézt nejlepší kombinaci, protože požadované vlastnosti nejsou vždy přítomny ve všech výstupech. (27) (36) (37)

- **Prompt:** queen with blue dress, crown, sitting on red throne, young, beautiful, powerfull, dark background, dslr, RAW photo, ultra quality, 8K UHD, high detailed, soft lighting, (skin pores, high detailed skin, sharp focus:1.2), film grain, Fujifilm XT3
- **Negative prompt:** (necklace:1.5), cropped, out of frame, duplicate, text, (painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render, deformed iris, deformed pupils:1.4), morbid, mutilated, extra fingers, fused fingers, too many fingers, mutated hands, poorly drawn hands, poorly drawn face, mutation, deformed, bad anatomy, bad proportions, extra limbs, cloned

face, cloned limbs, malformed limbs, missing arms, missing legs, extra arms, extra legs, ugly, long neck, disfigured, gross proportions, worst quality, low quality, blurry, hdr, jpeg artifacts, underexposed, overexposed



Obrázek 17 Realistická fotografie mocné královny s kompletním negativním promptu vytvořená ve Stable Diffusion

2.2.3 Váha klíčového slova

Váha klíčových slov ve Stable Diffusion představuje důležitý prvek, který ovlivňuje výsledky generovaných obrázků. Pouhé zahrnutí těchto slov do promptů samo o sobě nestačí k dosažení perfektních výsledků. Jednoduše řečeno, uživatel tímto generátoru sdělí, aby se důkladně zaměřil na specifické pojmy, zatímco by omezil svou pozornost na jiné. Tato možnost je zvláště užitečná, když se výstup blíží k požadovaným výsledkům, ale stále se liší v míře souladu s očekáváními.

Je také nutné správně nastavit poměr váhy mezi jednotlivými klíčovými slovy a určit jim prioritu. To lze upravit pomocí jednoduché syntaxe, která zahrnuje parametr „**(keyword:factor)**“. Hodnota faktoru určuje, jak podstatné je klíčové slovo v porovnání s ostatními. Pokud je faktor menší než 1, znamená to, že je méně důležité. Naopak, pokud je faktor větší než 1, naznačuje to, že je důležitější a mělo by získat větší váhu ve výsledném textu. Tato technika může být aplikována na klíčová slova týkající se objektu, stejně jako na všechny jeho kategorie, jako například styl a osvětlení.

Základní zřejmou aplikací je eliminace nežádoucích prvků z obrázků, tedy použití klíčového slova v negativním promptu. Pro představu je zvolena impozantní digitální

malba ulice v New Yorku, kde původně nebyla očekávána přítomnost automobilů. Následně tedy vznikne druhý obraz pomocí stejného počátečního promptu, jež určuje základní charakter celého díla, s přidáním negativního promptu „cars“, který provede vymazání většiny vozidel. Lze si ale všimnout, že na přiloženém obrázku jedno auto zůstalo. K dosažení lepšího výsledku, se doporučuje upravit negativní prompt na „(cars:1.3)“. To znamená, že klíčové slovo „cars“ je nyní považováno v rámci procesu difúze o 30 % důležitější. Na to poukazuje třetí a již finální výstup.

- **Prompt:** *street in New York, digital painting, autumn, vivid colors, high detailed, by Greg Rutkowski*
- **Negative prompt:** *(cars:1.3)*



Obrázek 18 Digitální malba ulice v New Yorku bez negativního promptu vytvořená ve Stable Diffusion



Obrázek 19 Digitální malba ulice v New Yorku s "cars" v negativním promptu vytvořená ve Stable Diffusion



Obrázek 20 Digitální malba ulice v New Yorku s "(cars:1.3)" v negativním promptu vytvořená ve Stable Diffusion

Nově vytvořené scény evidentně vykazují výraznou podobnost s původní verzí, avšak nejsou zcela identické. Pokud skutečně uživatel touží po původním provedení, je nezbytné použít **metodu inpaintingu**, která pečlivě odstraní přítomné rušivé objekty a současně zachová celistvost výjevu. Více je definována v kapitole 2.3.2.1.

V opačném smyslu, v rámci promptu, zvyšování váhy vykazuje tendenci generovat násobky objektů či vlastností definovaných klíčovým slovem. Redukce váhy zase projevuje sklon k vytváření menšího množství. Ovšem neplatí to vždy pro každý jednotlivý obrázek. (31) (33) (27) (36)

2.2.3.1 Syntaxe „()“ a „[]“

Způsob, jak upravit sílu klíčových slov, ekvivalentně spočívá v použití kulatých a hranatých závorek. Použití „(keyword)“ zvyšuje sílu klíčového slova o faktor 1,1, což je totožné s „(keyword:1.1)“. Naopak použití „[keyword]“ snižuje prioritu o faktor 0,9

shodující se s „(keyword:0.9)“. Vzhledem k multiplikativní povaze efektu je možné v této situaci kombinovat více operací uvnitř sebe, obdobně jako je tomu v oblasti algebry.

- | | |
|------------------------|------------------------|
| – (keyword) = 1,1 | – [keyword] = 0,9 |
| – ((keyword)) = 1,21 | – [[keyword]] = 0,81 |
| – (((keyword))) = 1,33 | – [[[keyword]]] = 0,73 |

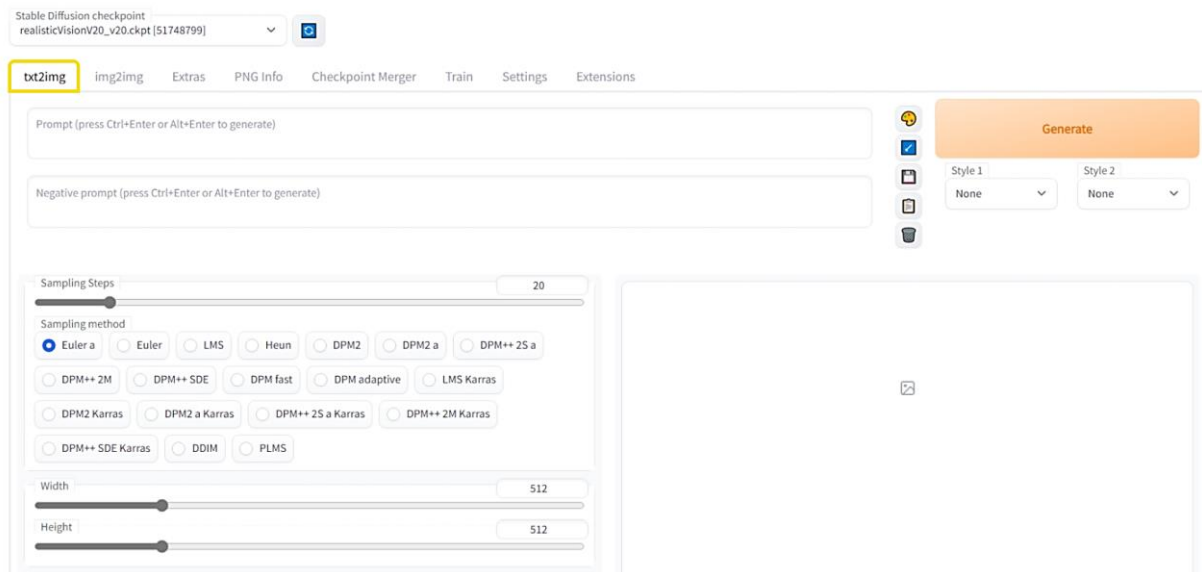
2.3 Technologie generování obrázků

Stable Diffusion umí vytvářet nové obrázky od základu pomocí textového zadání, které popisuje prvky, jenž mají být zahrnuty nebo vynechány ve výstupu. Dále umožňuje překreslení stávajících obrázků s novými prvky opět pomocí popisu (tzv. vedená syntéza obrazu) prostřednictvím mechanismu difúze a odšumění. Kromě toho model také umožňuje částečnou úpravu stávajících obrázků prostřednictvím dokreslování, pokud je použito vhodné uživatelské rozhraní, které tyto funkce podporuje. Existuje totiž mnoho různých otevřených implementací tohoto modelu.

Pomocí těchto technologií je možné vytvářet širokou škálu obrázků a grafiky, včetně realistických fotografií, designu, ilustrací, vizuálních efektů a podobně. (27)

2.3.1 Text-to-image

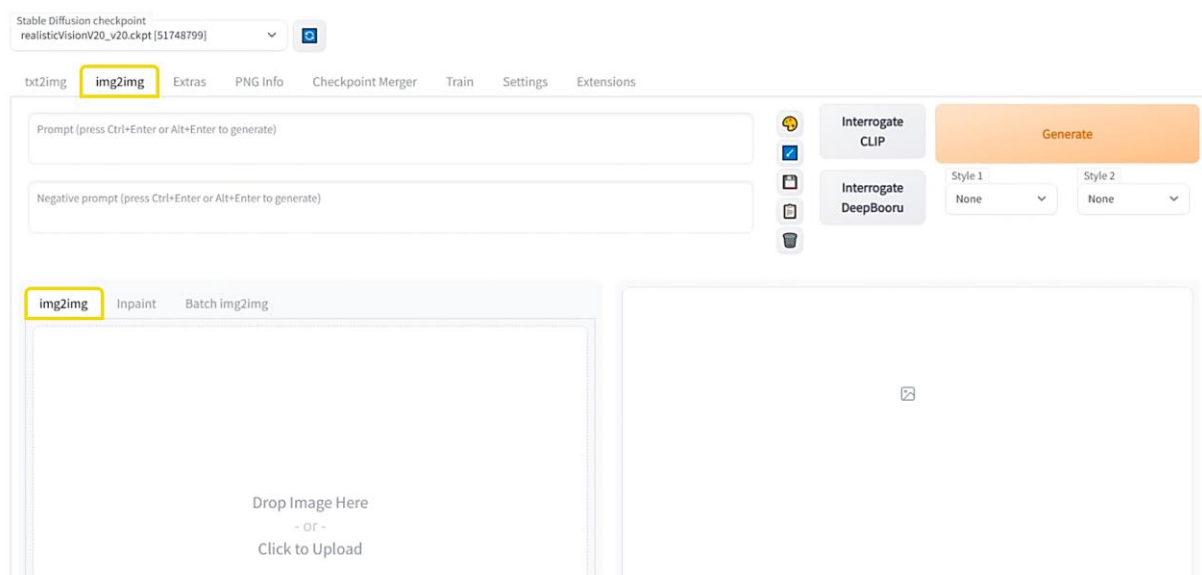
Samplovací skript text-to-image (txt2img) je výsledkem strojového učení, který dokáže generovat odpovídající obrázek **na základě textového popisu** v přirozeném jazyce. Tyto systémy se začaly vyvíjet od roku 2010 díky pokroku v oblasti hlubokých neuronových sítí. V podstatě se skládají z jazykového modelu převádějícího textový popis na latentní reprezentaci a generativního obrazového modelu, který na základě této reprezentace vytváří obraz. (27) (38) (39)



Obrázek 21 Okno technologie text-to-image

2.3.2 Image-to-image

Dalším výsledkem strojového učení je smplovací skript image-to-image (img2img), který využívá hluboký generativní model k syntéze obrázků **na základě textových popisů a obrázků**. Jeho hlavním cílem je naučit se převádět obrázky z jedné domény do druhé, a to s využitím bohatého prostoru transformací. Tento model dokáže provádět rozsáhlé spektrum úloh, jako je konverze černobílých obrazů na barevné, přeměna skic na fotorealistické, a dokonce i převod jednoho prostředí na jiné. Například může změnit zimní krajinu na letní nebo obrázek venkovské scenérie na městský pohled.



Obrázek 22 Okno technologie image-to-image

Jako ilustrativní příklad může posloužit fotografie jablka, která se díky využití algoritmu image-to-image proměnila v jedinečnou olejomalbu inspirovanou stylem malíře Vincenta van Gogha.

- **Prompt:** *red apple on wooden table, soft lighting, RAW photo*
- **Negative prompt:** *painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render, text*

Při aplikování technologie image-to-image byl prompt snadno upraven. Namísto „RAW photo“ byla zvolena fráze „oil painting“, která transformuje požadovaný obrázek. Doplněno bylo také jméno umělce textací „by Vincent van Gogh“, což dodává obrázku jeho typický styl. V případě negativního promptu bylo slovo „painting“ odstraněno, aby nebránilo vytvoření malby, a bylo nahrazeno slovy „RAW photo“ za účelem zabránění v tvorbě realistické fotografie. (40) (41)



Obrázek 23 Realistická fotografie vytvořená pomocí technologie text-to-image ve Stable Diffusion

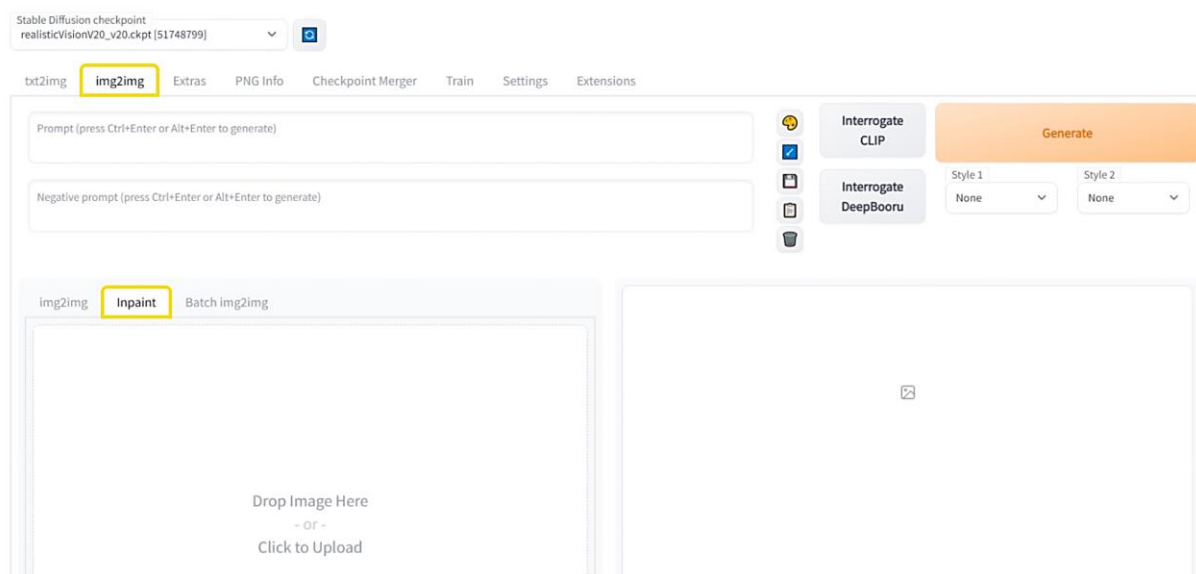


Obrázek 24 Olejomalba inspirovaná díly malíře Vincenta van Gogha vytvořená pomocí technologie image-to-image ve Stable Diffusion

2.3.2.1 Inpaint

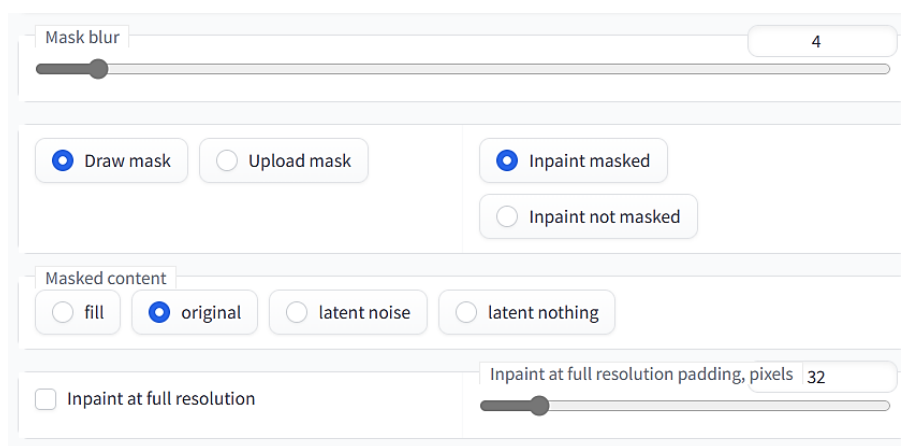
Flexibilita modelu Stable Diffusion poskytuje v rámci procesu image-to-image daleko rozsáhlejší schopnosti než pouhé generování obrazu.

Inpainting, známé také jako **dokreslování**, představuje schopnost úpravy obrázků, která uživatelům umožňuje modifikovat jednotlivé části a nahradit je jiným generovaným obsahem na základě textového vstupu. Model využívá informace z okolních pixelů k zajištění toho, aby generovaný obsah organicky zapadal do kontextu původního obrázku.



Obrázek 25 Okno technologie inpainting

Existuje mnoho nástrojů umožňujících změny v rámci samotného výstupu. Například pomocí „smazání“ určitých částí obrázku pro jejich zachování (inpaint masked) nebo aplikací masky na konkrétní oblast, která má projít změnou (inpaint not masked). Následně model vzniklý prostor vyplní novým obsahem. Mimo kreslení lze dokonce samostatně masku importovat. Funkce inpaintingu mohou být využity jak s reálnými obrázky ze skutečného světa, tak i s předem vygenerovanými.



Obrázek 26 Nástroje technologie inpainting

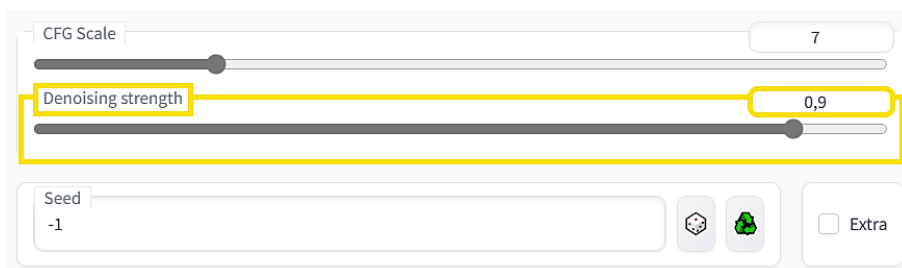
Nyní je představen postup, jak využít techniku dokreslování k opravě drobných defektů. Konkrétně dojde k odstranění automobilu z druhé digitální malby ulice v New Yorku, která je převzata z kapitoly 2.2.3 týkající se váhy klíčových slov. V tomto případě se díky inpaintingu zajistí, že celková scéna obrázku zůstane nedotčena, na rozdíl od pouhého zvýšení váhy slova „cars“ v negativním promptu, jež scénu patrně mění. Většinou stačí použít pouze prompt, který označuje, jak by upravený obrázek měl vypadat.

- **Prompt:** *empty street in New York, digital painting, autumn, vivid colors, high detailed, by Greg Rutkowski*



Obrázek 27 Aplikace masky na oblast vozidla pomocí režimu „inpaint masked“

Pokud nedochází ke kompletní zamýšlené změně, jako tomu je u tohoto obrázku, je nutné proces kombinovat se zintenzivněním „denoising strength“, v překladu odšumovací síly. To znamená, že bude generátor brát na prvotní obsah menší ohled a pravděpodobněji vygeneruje něco úplně nového. Původní hodnota proběhla změnou na výši 0,9. (42) (19) (43)



Obrázek 28 Posuvník hodnoty odšumovací síly



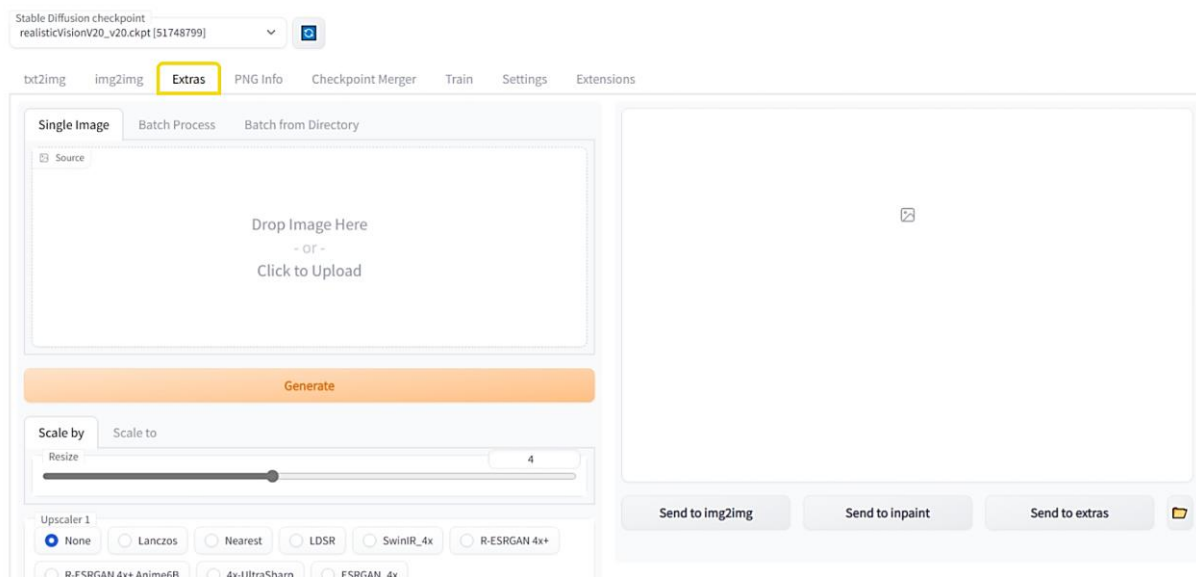
Obrázek 29 Digitální malba ulice v New Yorku před použitím inpaintingu vytvořená ve Stable Diffusion



Obrázek 30 Digitální malba ulice v New Yorku po použití inpaintingu vytvořená ve Stable Diffusion

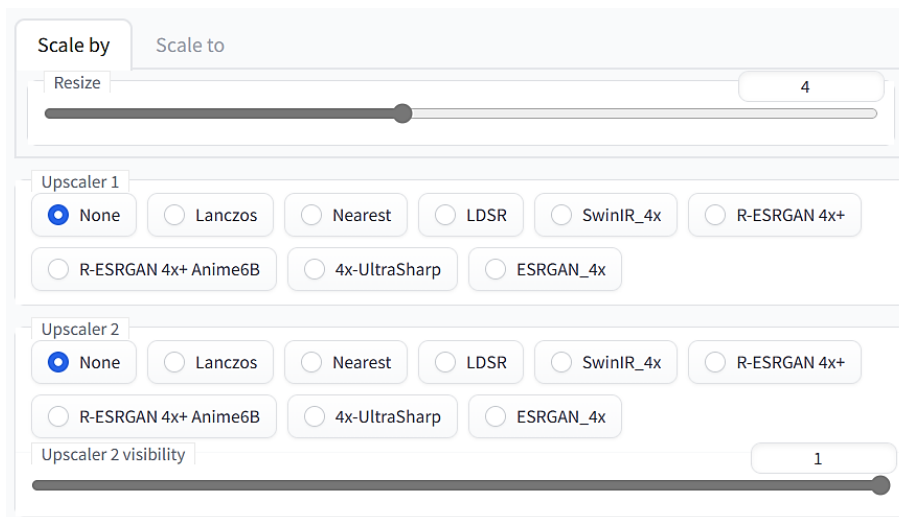
2.3.3 Extras

Sekce Extras nabízí uživatelům další možnosti pro práci s vygenerovanými obrázky. Pro zachování jejich maximální kvality má smysl je **upscalovat**, tedy zvětšit.



Obrázek 31 Okno sekce extras

Extras poskytuje výběr upscalovacího algoritmu, který upravuje velikost obrázků pomocí umělé inteligence různými způsoby. Uživatel si může zvolit násobek zvětšení až na hodnotu 8, anebo si přímo nastavit ve druhé záložce „Scale by“ počet pixelů, kterého by chtěl dosáhnout.



Obrázek 32 Rozhraní funkcí upscalingu a upscalovací algoritmy

U realistických fotografií podává nejlepší výsledky algoritmus SwinIR_4x nebo 4x-UltraSharp s nastavením čtyřnásobného zvětšení, jak naznačují jejich názvy. S drobnými nedostatky obstojí i ESRGAN_4x. Naopak u digitálních obrázků exceluje R-ESRGAN 4x+ Anime6B. Lze dokonce zkombinovat dva různé algoritmy a regulovat jejich viditelnost.

AI upscalery se liší od tradičních tím, že jsou trénovány na obrovském množství dat. Díky obsaženým znalostem jsou tedy schopny obnovit chybějící informace a zachovat detaily. (44) (45) (46)



Obrázek 33 Upscalovaný snímek lva



Obrázek 34 Detail upscalovaného snímku lva



Obrázek 35 Detail originálního snímku lva

2.4 Sampling steps

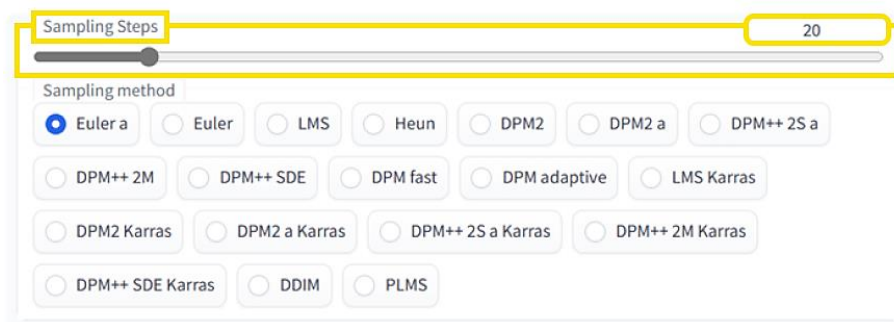
Samplovací kroky (vzorkování) se týkají počtu iterací, které model Stable Diffusion provede na základě daného promptu, aby přešel z náhodného Gaussova šumu na rozpoznatelný obraz. Znamená to tedy kolikrát bude obraz upraven, než uživatel obdrží finální výstup. Obecně platí, že čím vyšší je počet smplovacích kroků, tím podrobnější obrázek je za cenu delší doby zpracování. Určení optimálního počtu a požadovaného výstupního obrazu však závisí na konkrétní použité smplovací metodě.

Tento proces je demonstrován na níže uvedených sériích snímků.



Obrázek 36 Ukázka procesu generování obrazu za nižšího počtu smplovacích kroků ve Stable Diffusion

Výchozí hodnota ve webovém uživatelském rozhraní Stable Diffusion je 20, s minimem 1 a maximem 150.



Obrázek 37 Posuvník počtu samplovacích kroků

Na základě uvedených informací může uživatel zvolit vyšší počet samplovacích kroků s předpokladem, že to povede k většímu množství detailů a lepší kvalitě výsledného obrazu. Bohužel to však není vždy pravda. Tato volba by měla za následek velmi dlouhý čas zpracování, který vyžaduje výkonnější počítačové prostředky. Kromě toho, při určitém prahu přidáných detailů se dosáhne vrcholu a další vzorkování může ve skutečnosti zhoršit kvalitu obrazu spíše než ji zlepšit.



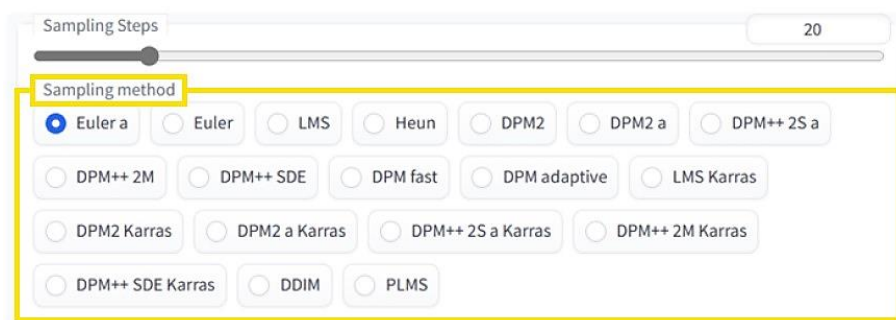
Obrázek 38 Ukázka procesu generování obrazu za vyššího počtu samplovacích kroků ve Stable Diffusion

Je zřejmé, že při stejném vstupním textu a všech ostatních parametrech model pokračuje v přidávání detailů do obrazu až do přibližně 80 samplovacích kroků. Nicméně s postupným zvyšováním jejich počtu začíná snímek ztrácet svou ostrost, což je výrazně patrné na textuře lví srsti, u digitálních obrázků ještě daleko více. Při dosažení 150 samplovacích kroků se také výrazně zvyšuje intenzita barev a kontrast. Obecně se tedy doporučuje pracovat s 60-80 kroky. (47) (48) (49)

2.5 Sampling method

Samplovací metody (vzorkovače) označují způsob, jakým jsou vykonávány jednotlivé smplovací kroky.

Ve Stable Diffusion web UI je celkově k dispozici 19 smplovacích metod. Až na pár výjimek, které fungují lépe či hůře, produkují většinou stejné obrázky s rostoucím počtem kroků.



Obrázek 39 Smplovací metody

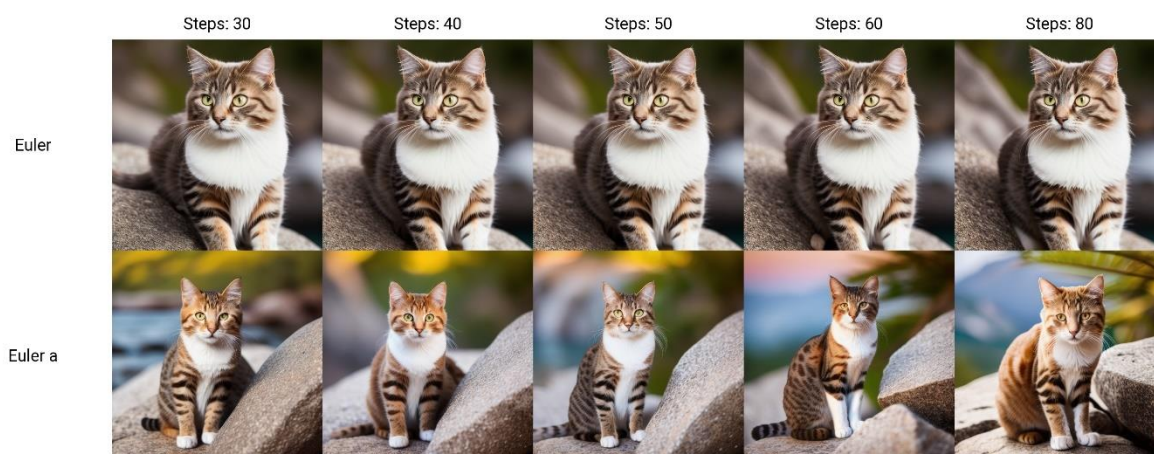
Nejde si nevšimnout, že názvy některých metod obsahují písmeno „a“. To znamená, že se jedná o „**vzorníky předků**“ (ancestral samplers). Nevýhodou jejich použití je, že zajišťují určitou náhodnost výstupu. Přidávají totiž do obrazu šum v každém jednotlivém kroku, což způsobuje nesbíhající se obraz. (49) (47)

2.5.1 Srovnání smplovacích metod

V porovnávacím testu bylo zjištěno, že se smplovací metoda **Euler** díky její univerzálnosti ukazuje jako nejlepší. Je to přímočarý vzorkovač, který se matematicky shoduje s Eulerovou metodou pro řešení obyčejných diferenciálních rovnic. Konverguje se a je zcela deterministický, což znamená, že při generování nedochází k přidání žádného náhodného šumu a obraz se po jednotlivých krocích sbíhá. Při navyšování počtu iterací tedy nedochází k tak velkým změnám, obsah na obrázku má stejné rozpoložení. Euler se považuje za nejjednodušší a mnoho dalších smplovacích metod sdílí jeho prvky.

Při tvorbě realistických snímků dopadl nejlépe předchůdce metody Euler, tedy **Euler a**. Vykazuje několik shodných rysů, avšak na každém kroku pracuje s velkým množstvím náhodného šumu. Kvůli tomu výsledný obrázek vypadá po každém generování odlišně. Nelze pomocí jej vykonávat malé úpravy.

Zde jsou prezentovány obrázky vytvořené metodou Euler společně s jejím předchůdcem Euler a.



Obrázek 40 Porovnání smplovací metody Euler s jejím předchůdcem Euler a

Metoda **Heun** představuje přesnější verzi Euler. Hlavní rozdíl spočívá v tom, že potřebuje předpovídat šum dvakrát v jednotlivém kroku, což způsobuje dvojnásobné zpomalení. Pro shodný výsledek s metodou Euler stačí snížit počet smplovacích kroků.



Obrázek 41 Obrázek s využitím smplovací metody Heun

LMS (Linear Multi-Step Method) se podobá rychlostí metodě Euler, ale je spíše vhodná pro generování „nefotografických“ výstupů. Díky své vysoké úrovni detailnosti vyniká při tvorbě kreseb, malířských děl a renderování. LMS má také verzi se speciálním odšumovacím plánem Karras. Realistické snímky z ní často působí nedotvořeně a mají příliš výrazné kontrasty (HDR). Je však třeba zdůraznit, že v případě krajinné fotografie výjimečně dosahuje dobrých výsledků.



Obrázek 42 Obrázky s využitím smplovacích metod LMS

Vzorkovače s předponou **DPM (Diffusion Probabilistic Model Solver)** patří do skupiny nově vyvinutých řešitelů difúzních modelů, které potřebují alespoň 10 kroků pro generování nezkresleného obrazu. Od sebe se liší pouze v několika nepatrných drobnostech. Přesnějším a pomalejším se zdá být DPM2, který u fotografií lidských postav výrazně vyhlazuje pleť. DPM++ přináší vylepšenou verzi původní implementace, zatímco DPM fast vykazuje nejhorší vlastnosti. DPM adaptive nerespektuje uživatelem zvolený počet kroků, přizpůsobuje si ho podle sebe. Může být pomalý a nezaručuje dokončení generování.



Obrázek 43 Obrázky s využitím smplovacích metod DPM

DDIM (Denoising Diffusion Implicit Model) a **PLMS (Pseudo Linear Multi-Step Method)** byly dodávány s původní verzí Stable Diffusion. Obecně se již nepoužívají, protože jsou zastaralé. PLMS je novější a rychlejší alternativa k metodě DDIM, která spadá mezi úplně první navržené pro difúzní modely.



Obrázek 44 Obrázky s využitím smplovacích metod DDIM a PLMS

2.6 Rozměry

Stable diffusion poskytuje prostřednictvím svého rozhraní možnost konfigurovat rozměry generovaného obrazu, což zahrnuje nastavení jak **šířky**, tak **výšky**.

Původní verze modelu byly natrénovány na datové sadě obsahující obrázky o **512 × 512 pixelů**. Tomuto rozlišení proto odpovídá i výchozí velikost generovaných výstupů. To znamená, že kvalita výrazně klesá, když uživatel odchyluje své požadavky od očekávaného rozlišení.

Aktualizace následně přinesla také funkční možnost 768 × 768 pixelů, což je v dnešním standardu stejně relativně nízké číslo. Pro srovnání, mnoho současných telefonů disponuje fotoaparáty, které zachycují snímky s minimálním rozlišením tisíců pixelů, zatímco jejich obrazovky jsou jen o něco menší. Snímek ze Stable Diffusion by tak na těchto zařízeních musel být uměle zvětšen, což by vedlo ke ztrátě kvality. Navíc komplexní scény generované modelem také často trpí nedostatečnou ostrostí, nepřesnostmi a nedokážou zachytit jemné detaily.

Pokud uživatel chce, aby konečná velikost obrázku byla 2048 × 2048, je lepší vygenerovat obrázek o velikosti 512 × 512 a pak jej 4x upscalovat. A to buď v rámci Stable Diffusion v Extras, nebo pomocí nějakého upscaleru třetí strany. (27) (45)

2.7 Batch count a batch size

Batch count je parametr určující, jak často se generuje sada obrázků. Při použití algoritmů zpracovávajících velké množství dat může být efektivní vytvářet dávky obrázků v pravidelných intervalech.

Na druhé straně, **batch size** ovlivňuje, kolik obrázků se generuje současně. Vysoká hodnota může vést k lepšímu výkonu, protože se zpracovává větší množství dat najednou. Je však důležité mít na paměti, že s rostoucím batch size se také zvyšují nároky na paměť VRAM.

Celkový počet vygenerovaných obrázků je pak dán násobením hodnoty batch size s počtem dávek (batch count). Tento výpočet umožňuje určit, kolik obrázků bude celkem generováno v rámci daného nastavení. (49)

3 Proces tvorby a optimalizace ilustrační grafiky

Proces tvorby a optimalizace ilustrační grafiky byl primárně zaměřen na tvorbu **fotorealistických snímků**, které mohou dále sloužit jako významný vizuální prvek ve všech odvětvích, od reklamy a marketingu po vzdělávání a zábavu.

Ne vždy je možné ve skutečnosti vytvořit fotografii, která by přesně odpovídala představám jednotlivce. Existují situace, kdy je obtížné nebo nemožné zajistit potřebné prostředí, subjekt nebo okolnosti pro pořízení požadovaného snímku. V takových případech však může být pomocí Stable Diffusion vytvořen jakýkoli obrázek, který tvůrce napadne.

Vzhledem k této potřebě byla vyvinuta metoda, která umožňuje generování obrázků na základě určených témat. Konkrétně byla vybrána čtyři témata: **člověk, zvíře, krajina a architektura**. Pro každé z nich byl vytvořen téměř univerzální textový vstup (prompt a negativní prompt), který byl seskupen do **znovupoužitelného stylu**, spolu se specifickými instrukcemi k dosažení co nejlepšího možného výstupu. V některých případech je potřeba text upravit, není totiž možné vytvořit zcela všeobecný postup generování.

Nejdůležitější je si uvědomit, že při vytváření realistických fotografií by měla být zakázána veškerá ostatní „nefotografická“ média. V negativním promptu by se proto vždy měla uplatňovat převaha termínů jako „*painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render*“ s vyšší váhou, například 1.4. Naopak v promptu by měla figurovat slova „*dslr, RAW photo, ultra quality, 8K UHD, high detailed, soft lighting, sharp focus, film grain, Fujifilm XT3*“ vyjadřující kvalitu výsledných snímků a způsob jejich pořízení. Byl zvolen fotoaparát Fujifilm XT3 společně s filmovým zrnem, protože podle několika studií podává ve Stable Diffusion nejlepší výsledky.

V neposlední řadě všechna z vybraných témat obsahují negativní prompt „*worst quality, low quality, blurry, jpeg artifacts, underexposed, overexposed*“ představují omezení špatné kvality a nepřesností, správná kompozice vznikne díky „*golden ratio, perfect composition, cropped, out of frame*“. Zákaz duplikací a různých vodoznaků, které obsahují na internetu většinou zpoplatněné nebo zvláště kvalitní fotografie, umožní přidání „*duplicate, text, signature, watermark*“

Cílený návod byl vypracován na základě dlouhodobého výzkumu, analýzy a experimentování s různými prompty. Zároveň bylo také zjištěno, že i když je obrázek

detailně popsán, nemusí vždy přesně odpovídat očekávání. Vždy je tedy nutné počítat s časem potřebným k testování, než se dosáhne optimálního výsledku.

Pro vytvoření vskutku realistických fotografií je podstatné využití nadstavbového checkpointu, který je speciálně vyškolen pro tuto úlohu. Mnoho uživatelů je neznalých jejich existence a omezuje se na používání základní verze Stable Diffusion v1.5. Existuje několik pokročilých modelů, avšak nejlépe hodnoceným je **Realistic Vision v2**, který je proto v celém procesu používán, včetně většiny ukázek v předešlých kapitolách.

- Realistic Vision v2
- Hassan blend 1.4
- F222
- Chillout Mix
- Dreamlike Photoreal
- URPM

3.1 Člověk

Stable Diffusion je široce využíván pro vytváření realistických portrétů lidí, které mohou být téměř nerozeznatelné od skutečných fotografií.

Jednou z největších komplikací je generování lidských končetin a obličejů kvůli nízké kvalitě dat v databázi LAION. Model jim nedokáže dostatečně porozumět z důvodu nedostatku reprezentativních prvků v datech, na kterých byl trénován. K dosažení vysoké kvality a realismu při generování takových obrázků může napomoci funkce „**face restoration**“ (obnovení obličeje), která dokáže opravit vady a deformace obličejů. Hlavním faktorem je zachycení detailů, textur a struktury lidské pokožky.

Pozor, téměř všechny checkpointy jsou náchylné ke generování explicitních obrázků. Pro jejich potlačení je tedy dobré použít výrazy týkající se oblečení, jako „*dressed*“ v promptu a „*naked*“ v negativním promptu.

Následně byl vytvořen z těchto poznatků styl s názvem **Realistic Human**.

- **Prompt:** *dslr, RAW photo, ultra quality, 8K UHD, high detailed, soft lighting, (skin pores, high detailed skin, sharp focus:1.2), film grain, Fujifilm XT3*
- **Negative prompt:** *cropped, out of frame, duplicate, text, (painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render, deformed iris, deformed pupils:1.4), morbid, mutilated, extra fingers, fused fingers, too many*

deformed, bad anatomy, bad proportions, extra limbs, cloned face, cloned limbs, malformed limbs, missing arms, missing legs, extra arms, extra legs, ugly, long neck, disfigured, gross proportions, worst quality, low quality, blurry, hdr, jpeg artifacts, underexposed, overexposed

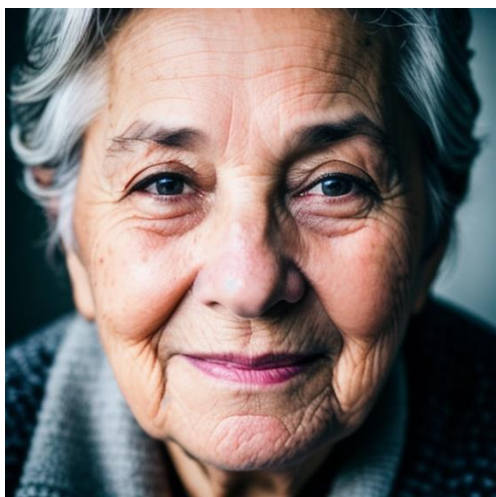
Za použití daného stylu postačuje jen do promptu napsat jednoduchá klíčová slova, která identifikují myšlenku uživatele – **hlavní postavu** a **atmosféru**. Doporučuje se zahrnout etnickou příslušnost, účes, outfit, aktivitu a okolí.



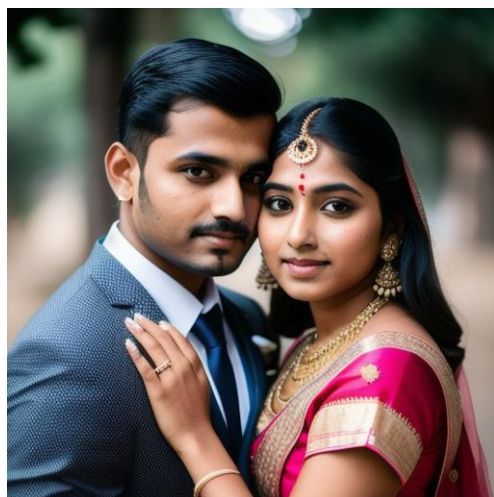
Obrázek 45 Ukázka obrázku dívky v Paříži vytvořeného pomocí stylu Realistic Human



Obrázek 47 Ukázka obrázku vojáka vytvořeného pomocí stylu Realistic Human



Obrázek 46 Ukázka obrázku staré ženy vytvořeného pomocí stylu Realistic Human



Obrázek 48 Ukázka obrázku páru vytvořeného pomocí stylu Realistic Human

3.2 Zvíře

Druhým nejoblíbenějším je generování zvířat, to je stejně fascinující jako generování lidí, přičemž oba procesy sdílejí mnoho podobných principů. Existují však drobné výjimky, které je třeba zohlednit.

Při generování zvířat je jedním z klíčových prvků jejich kožní struktura. Každý živočich má svůj specifický povrch těla, který se odlišuje od ostatních druhů. U některých zvířat je kůže hladká a lesklá, zatímco jiná mají šupiny, peří nebo srst.

Na světě se vyskytuje také skupina zvířat se specifickými „anomáliemi“, které vyžadují pečlivější přístup. Aby bylo dosaženo anatomicky správného výstupu, musí se rozdíly promítnout v promptu. Například sloni mají své charakteristické kly a chobot, které je třeba přesně vymodelovat. U jiných zvířat, zejména kopytníků, je důležité správně generovat tvar jejich končetin. Mořští živočichové zase mohou mít neobvyklé tvary těl, ploutve či dokonce chapadla.

Přidáním následujících frází do promptu se může zlepšit estetika snímků, především divoké zvěře – National Geographic Wildlife photo of the year, The American Landscape Contest, Wildlife photography contest.

Na základě podkladů byl sestaven speciální styl nazvaný **Realistic Animal**.

- **Prompt:** *dslr, RAW photo, ultra quality, 8K UHD, high detailed, soft lighting, (National Geographic Wildlife photo of the year, sharp focus, detailed eyes:1.2), film grain, Fujifilm XT3*
- **Negative prompt:** *cropped, out of frame, duplicate, (painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render, text, signature, watermark:1.4), (extra fingers, missing fingers, fused fingers, too many fingers, extra limbs, missing limbs, cloned limbs, malformed limbs, deformed iris, deformed pupils, deformed eyes:1.3), morbid, mutilated, mutation, deformed, bad anatomy, bad proportions, cloned face, duplicate tail, clone tail, deformed tail, deformed neck, duplicate neck, clone neck, duplicate hooves, clone hooves, deformed hooves, deformed trunk, disfigured, gross proportions, worst quality, low quality, blurry, hdr, jpeg artifacts, underexposed, overexposed, ugly*



Obrázek 49 Ukázka obrázku běžícího koně vytvořeného pomocí stylu Realistic Animal



Obrázek 51 Ukázka obrázku veverky vytvořeného pomocí stylu Realistic Animal



Obrázek 50 Ukázka obrázku motýla vytvořeného pomocí stylu Realistic Animal



Obrázek 52 Ukázka obrázku želvy v moři vytvořeného pomocí stylu Realistic Animal

3.3 Architektura

Stable Diffusion lze použít také ke generování architektury, umožňuje vytvářet fotorealistické snímky **interiérů** i **exteriérů**. Při psaní promptu je důležité specifikovat požadovaný architektonický sloh či umělecký styl. Jako příklad slouží secese, baroko, gotika, high-tech a mnoho dalších.

Pokud chce uživatel generovat fotografie vnitřních prostor, je vhodné zahrnout do promptu slovo „*interior*“. V případě, že se zaměří na tvorbu klasických venkovních staveb (domy, hrady, zámky, kostely, chrámy) měl by zdůraznit naopak „*exterior*“.

Pro dosažení nejvyšší kvality snímků může model využít přídatné informace jako „*artstation*“ nebo „*trending on artstation*“. ArtStation je totiž známá designerská webová

stránka, kde se často objevují vynikající rendery staveb. V promptu lze dále zmínit program OctaneRender, který je známý svou schopností vykreslovat realistické vizualizace architektury.

Pokud uživatel nemá speciální představu o křivkách nebo jiném zaobleném designu architektury, obvykle se přidává do promptu fráze „*straight lines*“ a do negativního „*wavy lines*“ s vyšší váhou. Model tak preferujete architekturu s přímými liniemi a ostrými tvary. Při fotografování staveb je důležité dodržovat správnou kompozici, to se dá najevo spojením slov „*perfect composition*“.

Níže je uveden styl **Realistic Architecture**, který lze uplatnit při generování architektury ve Stable Diffusion.

- **Prompt:** *dslr, RAW photo, ultra quality, 8K UHD, high detailed, perfect composition, perfect textures, trending on artstation, volumetric lighting, (award-winning photograph, sharp focus, octane render:1.2), film grain, Fujifilm XT3*
- **Negative prompt:** *cropped, out of frame, duplicate, text, (painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render, wavy lines:1.4), disfigured, deformed, worst quality, low quality, blurry, hdr, jpeg artifacts, underexposed, overexposed, ugly*



Obrázek 53 Ukázka obrázku moderního domu pomocí stylu Realistic Architecture



Obrázek 54 Ukázka obrázku futuristické v pomoci stylu Realistic Architecture



Obrázek 55 Ukázka obrázku interiéru pomocí stylu
Realistic Architecture



Obrázek 56 Ukázka obrázku kostela pomocí stylu
Realistic Architecture

3.4 Krajina

Generování krajiny je méně zajímavým a velmi jednoduchým tématem v rámci Stable Diffusion. Pro zlepšení výsledků je dobré zakombinovat do promptu konkrétní místo a jeho geografickou polohu.

Zvláště působivé snímky tvoří model se scénou zapadajícího slunce či za soumraku. Zvládne generování jakýchkoli obrázků přírody, což může zahrnovat pláž, opuštěný přístav, a dokonce i makro snímky rostlin. Zdůraznění klidu a krásy prostředí je klíčové.

Realistic Landscape představuje nejvíce objektivní a stručnou verzi stylu ze všech již zmíněných. Díky velké variabilitě vzhledu krajin na fotografiích by si měl uživatel textový vstup upravit dle svých představ. Pro dosažení vyššího stupně detailnosti a kontrastů ve výsledném obraze lze v promptu použít výraz „*hdr*“ nebo změnit smplovací metodu na LMS.

- **Prompt:** *dslr, RAW photo, ultra quality, 8K UHD, high detailed, (sharp focus:1.2), film grain, Fujifilm XT3*
- **Negative prompt:** *(painting, sketch, cartoon, drawing, anime, cgi, semi-realistic, 3d, render, text:1.4), worst quality, low quality, blurry, underexposed, overexposed, ugly*



Obrázek 57 Ukázka obrázku květinového pole na Novém Zélandu pomocí stylu Realistic Landscape



Obrázek 59 Ukázka obrázku řeky pomocí stylu Realistic Landscape



Obrázek 58 Ukázka obrázku zamrzlého moře na Islandu pomocí stylu Realistic Landscape



Obrázek 60 Ukázka obrázku krásné přírody pomocí stylu Realistic Landscape

Závěr

Práce byla zaměřena na generování fotorealistických snímků pomocí Stable Diffusion. Odkrývá rozsáhlé možnosti, které model přináší. I když zatím nebyl odhalen veškerý potenciál, je zřejmé, že se jedná o transformační technologii, která bude mít dalekosáhlý dopad na naši společnost.

V souladu se zadáním zahrnuje praktická část vytvoření sady obrázků s implementací klíčových slov, včetně porovnání smplovacích metod. Mým cílem bylo poskytnout komplexní pohled na generování ilustrací pomocí platformy Stable Diffusion, předat užitečné poznatky a doporučení pro začátečníky v této oblasti. Zkoumala jsem různé aspekty procesu tvorby a optimalizaci obrázků, včetně výběru vhodných promptů k navržení znovupoužitelného univerzálního stylu.

Lze konstatovat, že prostředí Stable Diffusion web UI poskytuje uživatelům mocný nástroj pro práci s rozsáhlými datovými sítěmi. Jeho uživatelsky přívětivé rozhraní, bohaté funkcionality a moderní technologické zázemí umožňují efektivně analyzovat textové vstupy a dosáhnout požadovaných výsledků.

Nicméně, navzdory dosaženému pokroku, existuje mnoho oblastí, ve kterých by mohla být práce dále rozvinuta. Jedním z těchto směrů je možnost rozšíření aplikací generovaných obrázků do oblasti umělecké tvorby. Umělci by mohli využívat tuto technologii k vytváření nových a inovativních děl, která by jinak byla obtížně dosažitelná. Kombinace lidské kreativity a výpočetních schopností vedou k neuvěřitelně zajímavým výtvarným výstupům.

Vývoj technologií představuje perspektivní cestu, která mění způsob, jakým vnímáme a interagujeme s vizuálním obsahem. Je tedy důležité pokračovat v objevování a inovacích, abychom plně využili potenciál generování obrázků a přispěli k budoucímu rozvoji naší společnosti. Lidé, kteří toho nevyužijí, riskují, že budou výrazně pozadu. Nejlepší čas pro začátek této cesty je právě teď a doufám, že tento průvodce bude sloužit jako pevný základ.

Seznam zkratk a odborných výrazů

AI

Artificial Intelligence (umělá inteligence) – oblast informatiky, která se zabývá vytvářením počítačových programů, které dokáží simulovat lidské chování a myšlení. (50)

Seznam obrázků

Obrázek 1 Prostředí Stable Diffusion web UI	6
Obrázek 2 Přehled Stable Diffusion checkpointů.....	7
Obrázek 3 Příkazový řádek pro prompt.....	9
Obrázek 4 Obrázek mocné královny vytvořený ve Stable Diffusion	10
Obrázek 5 Obrázek Emmy Watson jako mocné královny vytvořený ve Stable Diffusion	10
Obrázek 6 Olejomalba mocné královny vytvořená ve Stable Diffusion	11
Obrázek 7 Anime obrázek mocné královny vytvořený ve Stable Diffusion.....	11
Obrázek 8 Realistická fotografie mocné královny vytvořená ve Stable Diffusion	11
Obrázek 9 Fantazijní obrázek mocné královny v tématice kreslených Disney pohádek vytvořený ve Stable Diffusion	12
Obrázek 10 Olejomalba mocné královny inspirovaná díly malíře Leonarda da Vinciho vytvořená ve Stable Diffusion	13
Obrázek 11 Realistická fotografie mocné královny s měkkým osvětlením vytvořená ve Stable Diffusion	15
Obrázek 12 Realistická fotografie mocné královny s tmavým dramatickým osvětlením vytvořená ve Stable Diffusion	15
Obrázek 13 Příkazový řádek pro negativní prompt	16
Obrázek 14 Realistická fotografie mocné královny bez "without necklace" v promptu vytvořená ve Stable Diffusion	17
Obrázek 15 Realistická fotografie mocné královny s "without necklace" v promptu vytvořená ve Stable Diffusion	17
Obrázek 16 Realistická fotografie mocné královny s "necklace" v negativním promptu vytvořená ve Stable Diffusion	17
Obrázek 17 Realistická fotografie mocné královny s kompletním negativním promptu vytvořená ve Stable Diffusion	18
Obrázek 18 Digitální malba ulice v New Yorku bez negativního promptu vytvořená ve Stable Diffusion	19
Obrázek 19 Digitální malba ulice v New Yorku s "cars" v negativním promptu vytvořená ve Stable Diffusion.....	19
Obrázek 20 Digitální malba ulice v New Yorku s "(cars:1.3)" v negativním promptu vytvořená ve Stable Diffusion	19
Obrázek 21 Okno technologie text-to-image	21

Obrázek 22 Okno technologie image-to-image.....	21
Obrázek 23 Realistická fotografie vytvořená pomocí technologie text-to-image ve Stable Diffusion.....	22
Obrázek 24 Olejomalba inspirovaná díly malíře Vincenta van Gogha vytvořená pomocí technologie image-to-image ve Stable Diffusion.....	22
Obrázek 25 Okno technologie inpainting.....	23
Obrázek 26 Nástroje technologie inpainting.....	23
Obrázek 27 Aplikace masky na oblast vozidla pomocí režimu „inpaint masked“	24
Obrázek 28 Posuvník hodnoty odšumovací síly.....	24
Obrázek 29 Digitální malba ulice v New Yorku před použitím inpaintingu vytvořená ve Stable Diffusion	25
Obrázek 30 Digitální malba ulice v New Yorku po použití inpaintingu vytvořená ve Stable Diffusion	25
Obrázek 31 Okno sekce extras	25
Obrázek 32 Rozhraní funkcí upscalingu a upscalovací algoritmy	26
Obrázek 33 Upscalovaný snímek lva	26
Obrázek 34 Detail upscalovaného snímku lva	27
Obrázek 35 Detail originálního snímku lva.....	27
Obrázek 36 Ukázka procesu generování obrazu za nižšího počtu smplovacích kroků ve Stable Diffusion	27
Obrázek 37 Posuvník počtu smplovacích kroků.....	28
Obrázek 38 Ukázka procesu generování obrazu za vyššího počtu smplovacích kroků ve Stable Diffusion	28
Obrázek 39 Smplovací metody	29
Obrázek 40 Porovnání metody Euler s jejím předchůdcem Euler a.....	30
Obrázek 41 Obrázek s využitím smplovací metody Heun.....	30
Obrázek 42 Obrázky s využitím smplovacích metod LMS	31
Obrázek 43 Obrázky s využitím smplovacích metod DPM	31
Obrázek 44 Obrázky s využitím smplovacích metod DDIM a PLMS	32
Obrázek 45 Ukázka obrázku dívky v Paříži vytvořeného pomocí stylu Realistic Human	36
Obrázek 46 Ukázka obrázku staré ženy vytvořeného pomocí stylu Realistic Human	36
Obrázek 47 Ukázka obrázku vojáka vytvořeného pomocí stylu Realistic Human.....	36

Obrázek 48 Ukázka obrázku páru vytvořeného pomocí stylu Realistic Human.....	36
Obrázek 49 Ukázka obrázku běžícího koně vytvořeného pomocí stylu Realistic Animal	38
Obrázek 50 Ukázka obrázku motýla vytvořeného pomocí stylu Realistic Animal	38
Obrázek 51 Ukázka obrázku veverky vytvořeného pomocí stylu Realistic Animal	38
Obrázek 52 Ukázka obrázku želvy v moři vytvořeného pomocí stylu Realistic Animal...	38
Obrázek 53 Ukázka obrázku moderního domu pomocí stylu Realistic Architecture	39
Obrázek 54 Ukázka obrázku futuristické v pomocí stylu Realistic Architecture	39
Obrázek 55 Ukázka obrázku interiéru pomocí stylu Realistic Architecture	40
Obrázek 56 Ukázka obrázku kostela pomocí stylu Realistic Architecture.....	40
Obrázek 57 Ukázka obrázku květinového pole na Novém Zélandu pomocí stylu Realistic Landscape.....	41
Obrázek 58 Ukázka obrázku zamrzlého moře na Islandu pomocí stylu Realistic Landscape.....	41
Obrázek 59 Ukázka obrázku řeky pomocí stylu Realistic Landscape	41
Obrázek 60 Ukázka obrázku krásné přírody pomocí stylu Realistic Landscape	41

Použité zdroje

1. **IT slovník.** Co je to Neuronová síť? *IT slovník*. [Online] [Citace: 9. Duben 2023.]
<https://it-slovník.cz/pojem/neuronova-sit>.
2. **Wikimedia, přispěvatelé projektů.** Umělá neuronová síť. *Wikipedie: otevřená encyklopedie*. [Online] 5. Březen 2023. [Citace: 9. Duben 2023.]
https://cs.wikipedia.org/wiki/Um%C4%9Bl%C3%A1_neuronov%C3%A1_s%C3%AD%C5%A5.
3. **Joe Comp.** Co je hluboké učení a neuronová síť. *Joe Comp.* [Online] [Citace: 9. Duben 2023.] <https://cs.joecomp.com/what-is-deep-learning-and-neural-network>.
4. **IBM.** What are neural networks? *IBM*. [Online] [Citace: 9. Duben 2023.]
<https://www.ibm.com/topics/neural-networks>.
5. **Slater, Neil.** Can people use neural networks without providing the set of training data? *Artificial Intelligence Stack Exchange*. [Online] 5. Červenec 2021. [Citace: 22. Květen 2023.] <https://ai.stackexchange.com/questions/28524/can-people-use-neural-networks-without-providing-the-set-of-training-data>.
6. **SAP Insights.** Co je to strojové učení? *SAP*. [Online] [Citace: 19. Duben 2023.]
<https://www.sap.com/cz/products/artificial-intelligence/what-is-machine-learning.html>.
7. **Matoušek, Václav.** Strojové učení. *KIV*. [Online] 29. Listopad 2022. [Citace: 22. Květen 2023.] https://www.kiv.zcu.cz/studies/predmety/uzi/Folie_ZS/Stroj_uceni.pdf.
8. **Wikimedia, přispěvatelé projektů.** Strojové učení. *Wikipedie: otevřená encyklopedie*. [Online] 5. Březen 2023. [Citace: 22. Květen 2023.]
https://cs.wikipedia.org/wiki/Strojov%C3%A9_u%C4%8Den%C3%AD.
9. **Branscomb, Mary.** Strojové učení ve firemním IT. *Computerworld*. [Online] 30. Prosinec 2018. [Citace: 11. Květen 2023.]
<https://www.computerworld.cz/clanky/strojove-uceni-ve-firemnim-it/>.
10. **Wikimedia, přispěvatelé projektů.** Učení s učitelem. *Wikipedie: otevřená encyklopedie*. [Online] 6. Březen 2023. [Citace: 11. Květen 2023.]
https://cs.wikipedia.org/wiki/U%C4%8Den%C3%AD_s_u%C4%8Ditelem.

11. **Kod'ousková, Barbora.** Co je strojové učení a jak souvisí s umělou inteligencí? *Rascasone*. [Online] 13. Duben 2021. [Citace: 11. Květen 2023.]
<https://www.rascasone.com/cs/blog/strojove-uceni-ml-metody-klasifikace>.
12. **Šťastný, Petr.** Rozpoznávání objektů pomocí neuronových sítí. *Informační systém MUNI*. [Online] 2014. [Citace: 22. Květen 2023.]
https://is.muni.cz/th/rr91b/Stastny_Petr.pdf.
13. **OpenAI.** CLIP. *GitHub*. [Online] 20. Únor 2023. [Citace: 22. Květen 2023.]
<https://github.com/openai/CLIP>.
14. **Gwanghyun Kim, Taesung Kwon, Jong Chul Ye.** DiffusionCLIP: Text-Guided Diffusion Models for Robust Image Manipulation. *Arxiv*. [Online] 11. Srpen 2022. [Citace: 22. Květen 2023.] <https://arxiv.org/abs/2110.02711>.
15. **Alec Radford, Ilya Sutskever, Jong Wook Kim, Gretchen Krueger, Sandhini Agarwal.** CLIP: Connecting text and images. *OpenAI*. [Online] 5. Leden 2021. [Citace: 22. Květen 2023.] <https://openai.com/research/clip/>.
16. **O'Connor, Ryan.** Introduction to Diffusion Models for Machine Learning. *AssemblyAI*. [Online] 12. Květen 2022. [Citace: 23. Květen 2023.]
<https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/>.
17. **AI dětem, z.s.** Slovníček. *AI dětem*. [Online] [Citace: 23. Květen 2023.]
<https://aidetem.cz/slovnicek-pojmu-umele-inteligence/>.
18. **Vaněk, Karel.** Obrázky generované AI - K čemu jsou dobré a o co jde? *TechFeed*. [Online] 26. Říjen 2022. [Citace: 20. Květen 2023.]
<https://techfeed.cz/porad/techfeed/obrazky-generovane-ai-k-cemu-jsou-dobre-a-o-co-jde>.
19. **Vivek Muppalla, Sean Hendryx.** Diffusion Models: A Practical Guide. *Scale*. [Online] 19. Říjen 2022. [Citace: 20. Květen 2023.] <https://scale.com/guides/diffusion-models-guide>.
20. **Yalalov, Damir.** ChatGPT dokáže řešit jednoduché úlohy strojového učení, jako je klasifikace a kategorizace. *Metaversse Post*. [Online] 3. Leden 2023. [Citace: 20. Květen 2023.] <https://mpost.io/cs/chatgpt-can-solve-simple-machine-learning-tasks-as-classification-and-categorization/>.

21. **OpenAI.** DALL·E 2. *OpenAI*. [Online] [Citace: 23. Květen 2023.]
<https://openai.com/product/dall-e-2>.
22. **Javůrek, Karel.** DALL-E 2 se otevírá všem. Zkuste AI, která vám nakreslí, na co si vzpomenete. *Živě*. [Online] 29. Září 2022. [Citace: 23. Květen 2023.]
<https://www.zive.cz/clanky/vyzkouseli-jsme-dall-e-2-nocni-muru-vsech-grafiku-umela-intelligence-zvladne-narocne-kolaze-a-ma-smysl-pro-humor/sc-3-a-217633/default.aspx>.
23. **Holt, Kris.** Google claims its text-to-image AI delivers 'unprecedented photorealism'. *Engadget*. [Online] 24. Květen 2022. [Citace: 23. Květen 2023.]
https://www.engadget.com/google-imagen-text-to-image-ai-unprecedented-photorealism-144205123.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuYmluZy5jb20v&guce_referrer_sig=AQAAAEgyKV57Yd_u3bMQjpYtMqbmnhNqZsY9KaFqPzSSymmClimAs6NBEYxQeGi2Hco4JhpSDnUibtC4a2.
24. **Mae, Meta.** Stable Diffusion vs Midjourney vs DALL·E 2. *Buildspace*. [Online] 24. Leden 2023. [Citace: 23. Květen 2023.] <https://buildspace.so/notes/stable-diffusion-vs-midjourney-vs-dalle-2>.
25. **Wikimedia, přispěvatelé projektů.** Midjourney. *Wikipedie: otevřená encyklopedie*. [Online] 14. Duben 2023. [Citace: 24. Květen 2023.]
<https://cs.wikipedia.org/wiki/Midjourney>.
26. **Kamph, Sebastian.** Don't make these 7 mistakes in Stable diffusion. *YouTube*. [Online] 20. Listopad 2022. [Citace: 24. Květen 2023.]
<https://www.youtube.com/watch?v=b8xWjrztAPY>.
27. **Wikimedia, contributors.** Stable Diffusion. *Wikipedia: the free encyclopedia*. [Online] 23. Květen 2023. [Citace: 24. Květen 2023.]
https://en.wikipedia.org/wiki/Stable_Diffusion.
28. **AUTOMATIC1111.** Stable Diffusion web UI. *GitHub*. [Online] 23. Květen 2023. [Citace: 27. Květen 2023.] <https://github.com/AUTOMATIC1111/stable-diffusion-webui>.
29. **Ma, Yubin.** The 10 Best Stable Diffusion Models by Popularity (SD Models Explained). *Aituts*. [Online] 22. Květen 2023. [Citace: 29. Květen 2023.]
<https://aituts.com/models/>.

30. **Robin Rombach, Patrick Esser.** Stable Diffusion. *Hugging Face*. [Online] Červen 2022. [Citace: 29. Květen 2023.] <https://huggingface.co/CompVis/stable-diffusion>.
31. **Andrew.** Stable Diffusion prompt: a definitive guide. *Stable Diffusion Art*. [Online] 14. Květen 2023. [Citace: 24. Květen 2023.] <https://stable-diffusion-art.com/prompt-guide/>.
32. **Brown, Naomi.** How to Write Amazing Stable Diffusion Prompts: A Definitive Guide for You. *Fotor*. [Online] 22. Květen 2023. [Citace: 24. Květen 2023.] <https://www.fotor.com/blog/stable-diffusion-prompts/>.
33. **Lewis, Nick.** How to Write an Awesome Stable Diffusion Prompt. *How-To Geek*. [Online] 23. Prosinec 2022. [Citace: 24. Květen 2023.] <https://www.howtogeek.com/833169/how-to-write-an-awesome-stable-diffusion-prompt/>.
34. **Heikkiläarchive, Melissa.** Tento umělec dominuje umění generovanému umělou inteligencí. A není z toho nadšený. *MIT Technology Review*. [Online] 16. Září 2022. [Citace: 25. Květen 2023.] <https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>.
35. **Kapoor, Mukund.** 40 Ai Prompts To Generate High-Quality Photography Portraits. *Great Ai Prompts*. [Online] 2023. [Citace: 25. Květen 2023.] <https://www.greataiprompts.com/imageprompt/ai-prompts-for-high-quality-realistic-photography-portraits/>.
36. **Andrew.** How to use negative prompts? *Stable Diffusion Art*. [Online] 6. Únor 2023. [Citace: 27. Květen 2023.] <https://stable-diffusion-art.com/how-to-use-negative-prompts/>.
37. —. How does negative prompt work? *Stable Diffusion Art*. [Online] 8. Únor 2023. [Citace: 27. Květen 2023.] <https://stable-diffusion-art.com/how-negative-prompt-work/>.
38. **Wikimedia, contributors.** Text-to-image model. *Wikipedia: the free encyclopedia*. [Online] 19. Květen 2023. [Citace: 24. Květen 2023.] https://en.wikipedia.org/wiki/Text-to-image_model.

39. **Hugging Face.** Text-to-Image Generation. *Hugging Face*. [Online] [Citace: 24. Květen 2023.]
https://huggingface.co/docs/diffusers/api/pipelines/stable_diffusion/text2img.
40. **Yaxing Wang, Hector Laria Mantecón, Joost van de Weijer, Laura Lopez-Fuentes, Bogdan Raducanu.** TransferI2I: Transfer Learning for Image-to-Image Translation from Small. *Arxiv*. [Online] 14. Květen 2021. [Citace: 24. Květen 2023.] <https://arxiv.org/pdf/2105.06219.pdf>.
41. **McQuade, Mark.** How to Build a Stable Diffusion Image-to-Image Pipeline. *Roboflow*. [Online] 10. Únor 2023. [Citace: 24. Květen 2023.]
<https://blog.roboflow.com/stable-diffusion-image-to-image-pipeline/>.
42. **Andrew.** Beginner's guide to inpainting (step-by-step examples). *Stable Diffusion Art*. [Online] 26. Únor 2023. [Citace: 26. Květen 2023.] https://stable-diffusion-art.com/inpainting_basics/.
43. **Wikimedia, contributors.** Inpainting. *Wikipedia: the free encyclopedia*. [Online] 3. Květen 2023. [Citace: 28. Květen 2023.]
<https://en.wikipedia.org/wiki/Inpainting>.
44. **Binks.** HOW TO MASTER AI UPSCALING/UP-RES | Stable Diffusion Web UI. *YouTube*. [Online] 21. Leden 2023. [Citace: 29. Květen 2023.]
<https://www.youtube.com/watch?v=fjWcRU7Ue-0>.
45. **Andrew.** How to use AI upscaler to improve image details. *Stable Diffusion Art*. [Online] 12. Květen 2023. [Citace: 29. Květen 2023.] https://stable-diffusion-art.com/ai-upscaler/#How_does_AI_upscaler_work.
46. **AUTOMATIC1111.** Features. *GitHub*. [Online] 9. Květen 2023. [Citace: 29. Květen 2023.] <https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features>.
47. **Andrew.** Stable Diffusion Samplers: A Comprehensive Guide. *Stable Diffusion Art*. [Online] 2. Duben 2023. [Citace: 28. Květen 2023.] <https://stable-diffusion-art.com/samplers/>.
48. **Inglewood, Alex.** Guide: What are Sampling Steps and How To Reduce Them in Stable Diffusion. *Once upon an algorithm*. [Online] 19. Březen 2023. [Citace: 28.

Květen 2023.] <https://onceuponanalgorithm.org/guide-what-are-sampling-steps-and-how-to-reduce-them-in-stable-diffusion/>.

49. **InstallGentoo Wiki, contributors.** Stable Diffusion. *InstallGentoo Wiki*. [Online] 22.

Říjen 2022. [Citace: 28. Květen 2023.]

https://wiki.installgentoo.com/wiki/Stable_Diffusion#Generation_Parameters.

50. **Wikimedia, přispěvatelé projektů.** Umělá inteligence. *Wikipedie: otevřená*

encyklopedie. [Online] 30. Květen 2023. [Citace: 30. Květen 2023.]

https://cs.wikipedia.org/wiki/Um%C4%9Bl%C3%A1_inteligence.

A. Seznam příložených souborů

Na přiloženém datovém nosiči se nacházejí následující soubory a složky:

- **RP2023-Ryšavá-Natálie-L3-Využití_Stable_Diffusion_pro_vytváření_ilustrační_grafiky.docx**
– editovatelná verze dokumentace ročníkové práce
- **RP2023-Ryšavá-Natálie-L3-Využití_Stable_Diffusion_pro_vytváření_ilustrační_grafiky.pdf**
– tisknutelná verze dokumentace ročníkové práce
- **Styles.csv** – kompletní dokumentace stylů
- **KatalogObrázků** – kompletní dokumentace obrázků
- **Extras** – složka s doplňujícími obrázky