Univerzita Karlova

Pedagogická fakulta

Katedra informačních technologií a technické výuky

BAKALÁŘSKÁ PRÁCE

Multimediální průvodce generováním obrázků prostřednictvím lokálně provozovaného modelu umělé inteligence

Multimedia guide to image generation through locally operated artificial intelligence model

Lucie Drabčíková

Vedoucí práce:	PhDr. Jiří Leipert, Ph.D.
Studijní program:	Informační technologie se zaměřením na vzdělávání (B0114A140004)
Studijní obor:	B IT 20 (0114RA140004)

2024

Odevzdáním této bakalářské práce na téma Multimediální průvodce generováním obrázků prostřednictvím lokálně provozovaného modelu umělé inteligence potvrzuji, že jsem ji vypracovala pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále potvrzuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

V Praze 02.12.2024

Ráda bych poděkovala PhDr. Jiří Leipert, Ph.D. za cenné rady, věcné připomínky, motivaci a vstřícnost při konzultacích mé bakalářské práce. Mé poděkováni patři též Jaroslavu Mikulíkovi za spolupráci při získávání údajů pro technicky zaměřené části práce. Děkuji také své rodině za podporu při studiu.

ABSTRAKT

Tato bakalářská práce se zaměřuje na vytvoření multimediálního průvodce pro generování obrázků pomocí lokálně provozovaného modelu umělé inteligence Stable Diffusion. Práce je rozdělena na teoretickou a praktickou část. V teoretické části jsou shrnuty základní principy a technologie umělé inteligence používané při generování obrázků, včetně srovnání cloudových a lokálně provozovaných nástrojů. Důraz je kladen na software Stable Diffusion s instalátorem a ovládacím rozhraním EasyDiffusion. Praktická část zahrnuje vytvoření webových stránek s videonávody, které uživatele provedou instalací a používáním Stable Diffusion a EasyDiffusion. Výsledkem je ucelený průvodce, který uživatelům, zejména studentům vysokých škol, umožní efektivně využívat technologii generování obrázků na lokálních zařízeních. Průvodce je navržen tak, aby krok za krokem usnadnil uživatelům práci s těmito nástroji a přiblížil jim fungování umělé inteligence v oblasti tvorby grafiky.

KLÍČOVÁ SLOVA

Umělá inteligence, generování obrázků, grafika, Stable Diffusion, EasyDiffusion, videoprůvodce, hluboké učení

ABSTRACT

This work focuses on creating a multimedia guide for generating images using the locally operated artificial intelligence model Stable Diffusion. The work is divided into theoretical and practical parts. The theoretical part summarizes the basic principles and technologies of artificial intelligence used in image generation, including a comparison of cloud-based and locally operated tools. Emphasis is placed on the Stable Diffusion software with the EasyDiffusion installer and user interface. The practical part involves creating a website with video tutorials that guide users through the installation and use of Stable Diffusion and EasyDiffusion. The result is a comprehensive guide that enables users, especially university students, to effectively utilize image generation technology on local devices. The guide is designed to facilitate users' work with these tools step by step and to introduce them to the functioning of artificial intelligence in the field of graphic creation.

KEYWORDS

Artificial intelligence, image generation, graphics, Stable Diffusion, EasyDiffusion, video tutorial, deep learning

Obsah

Úvod ... 8

Cíle teoretické části ... 8

Cíle praktické části ... 8

Část I – teoretická část ... 9

1 Definování klíčových pojmů inteligence pro generování obrázků ... 10

1.1 Umělá inteligence (AI) ... 10

1.2 Neuronová síť ... 10

1.3 Strojové učení ... 10

1.4 Hluboké učení ... 11

1.5 Generativní umělá inteligence ... 11

1.6 Difúzní model ... 11

1.7 Model latentní difúze ... 12

1.8 Prompt ... 12

1.9 Datová sada ... 12

2 Stable diffusion a instalátor EasyDiffusion ... 13

2.1 Výhody lokálního běhu AI oproti běhu v cloudu ... 13

2.1.1 Soukromí a bezpečnost dat ... 13

2.1.2 Nezávislost na internetovém připojení ... 13

2.1.3 Výkon a rychlost ... 14

2.1.4 Cena ... 14

2.1.5 Možnosti generování ... 14

2.2 Nevýhody oproti online prostředí ... 16

2.2.1 Požadavky na hardware ... 16

2.2.2 Instalace a údržba ... 16

2.3 Parametry generování obrázku....17
2.3.1 Prompt (výzva)....17
2.3.2 Image Modifiers (modifikátory obrázku)....19
2.3.3 Seed (semínko)....19
2.3.4 Number of Images (počet obrázků)....20
2.3.5 Model....20
2.3.6 ControlNet....29
2.3.7 VAE (Variational Autoencoder)....30
2.3.8 Sampler (vzorkovač)....31
2.3.9 Image Size (velikost obrázku)....31
2.3.10 Inference Steps (počet kroků)....31
2.3.11 Guidance Scale (naváděcí škála)....32
2.3.12 Seamless Tiling (bezešvé dlaždice)....32
2.3.13 Output Format (výstupní formát)....32
2.3.14 Přepnutí akcelerovaného běhu na běh na procesoru....33
2.4 Text to image a image to image metody generování obrázku....33
3 Hardwarové požadavky na provoz AI Stable Diffusion....34
3.1 Provoz softwarově na procesoru....34
3.2 Akcelerace pomocí GPU....34
3.3 Akcelerace pomocí specializovaných čipů....35
3.4 Proč je běh umělé inteligence na GPU nebo AI akcelerátorech rychlejší....35
3.5 Porovnání řešení vzhledem k výkonu....37
3.5.1 Testovací PC sestava 1....37
3.5.2 Testovací sestava 2....38
3.5.3 Testovací sestava 3....38

Část II – praktická část....41
4 Úvod praktické části....42
4.1 Postup práce....42
4.1.1 Výběr tématu....42
4.1.2 Vyhledání a studium zdrojů....43
4.1.3 Analýza principu generování obrázků....43
4.1.4 Položení teoretických základů....43
4.1.5 Sestavení scénářů a kompletace videonávodů....43
4.1.6 Navržení struktury webu a jeho realizace....43
4.2 Hlavní cíle a funkce webových stránek....44
4.2.1 Seznámení s pojmy....44
4.2.2 Videoprůvodci....45
4.2.3 Galerie praktických ukázek....46
4.2.4 Více o problematice generování....47
4.3 Tvorba webu....48
4.4 Tvorba videoprůvodců....49
4.4.1 Jak nainstalovat Stable Diffusion pomocí EasyDiffusion....50
4.4.2 Princip generování obrázků ve Stable Diffusion....52
4.4.3 Jak na prompty pro Stable Diffusion....55
4.4.4 Přehled a použití modelů v Stable Diffusion....61
Závěr....63
Seznam použitých informačních zdrojů....64

Úvod

Generování obrázků pomocí umělé inteligence je dnes jednou z nejvíce se rozvíjejících oblastí strojového učení. Tato práce byla motivována potřebou přiblížit tuto technologii nejen studentům.

V současné době existuje několik metod generování obrázků, které zahrnují cloudové i lokálně provozované nástroje, z nichž každá má své výhody a nevýhody. Cloudová řešení často nabízejí vyšší výpočetní výkon a snadný přístup, zatímco lokální nástroje, jako je Stable Diffusion s rozhraním EasyDiffusion, poskytují větší kontrolu nad daty a procesem generování.

Cílem je vytvořit komplexní multimediální průvodce, který uživatelům poskytne potřebné znalosti a dovednosti pro efektivní práci s modelem Stable Diffusion za použití instalátoru EasyDiffusion. Práce bude rozdělena na teoretickou a praktickou část, které společně poskytnou ucelený obraz o možnostech a využití těchto technologií

Cíle teoretické části

- Představit základní principy umělé inteligence pro generování obrázků: Vysvětlit klíčové pojmy a technologie, jako je hluboké učení, neuronové sítě a generativní modely.
- Popis procesu generování obrázků: detaily o tom, jak AI generuje obrázky z textových popisů pomocí difúzních modelů.
- Srovnání cloudových a lokálních nástrojů: analyzovat výhody i nevýhody používání lokálně provozovaných modelů oproti cloudovým řešením.
- Představení nástrojů Stable Diffusion a EasyDiffusion: poskytnout přehled a výhody těchto nástrojů v kontextu generování obrázků.

Cíle praktické části

- Vytvořit webové stránky s videonávody: Vytvořit multimediálního průvodce generováním obrázků prostřednictvím lokálně provozovaného modelu umělé inteligence pro studenty vysokých škol.

Část I – teoretická část

1 Definování klíčových pojmů inteligence pro generování obrázků

Umělá inteligence (z anglického artificial intelligence, zkráceně AI) představuje široký obor informatiky, který se zaměřuje na vytváření systémů schopných vykonávat úkoly, jež by běžně vyžadovaly lidskou inteligenci. V kontextu generování obrázků se AI využívá k vytváření nových, realistických či stylizovaných vizuálů na základě specifických vstupů. Tyto technologie jsou založeny na pokročilých algoritmech strojového a hlubokého učení.

Pro lepší porozumění tématu generování obrázků pomocí umělé inteligence je nezbytné seznámit se s klíčovými pojmy, které se v této oblasti používají.

1.1 Umělá inteligence (AI)

Umělá inteligence je technologie, která umožňuje počítačům a strojům napodobovat lidskou inteligenci a schopnost řešit problémy. AI, jako významná oblast informatiky, zahrnuje a často se spojuje se strojovým učením a hlubokým učením. Tato technologie se využívá k tomu, aby stroje dokázaly vykonávat úkoly, které by běžně vyžadovaly lidskou schopnost myšlení, učení a rozhodování [1].

1.2 Neuronová síť

Neuronová síť je výpočetní model inspirovaný strukturou lidského mozku. Skládá se z vrstev propojených neuronů, které spolu komunikují pomocí vážených spojů. Každý neuron přijímá vstupy, zpracovává je a předává je dalším neuronům v síti. Neuronové sítě jsou základním stavebním kamenem pro AI algoritmy, včetně těch používaných pro generování obrázků [2] [3].

1.3 Strojové učení

Strojové učení je podmnožinou AI, která umožňuje systémům učit se a zlepšovat na základě zkušeností. Modely strojového učení jsou trénovány na velkém množství dat a používají algoritmy, které se přizpůsobují a zlepšují své výkony při řešení konkrétních úkolů. Vzhledem k požadavkům a datům je možné použít jeden ze čtyř modelů: bez dohledu, s dohledem, s polodohledem nebo posilování. V kontextu generování obrázků se strojové učení používá k tomu, aby se modely naučily vztahy mezi textovými popisy a vizuálními prvky obrazů [2] [4].

1.4 Hluboké učení

Hluboké učení je specializovaná oblast strojového učení, která využívá vícevrstvé neuronové sítě k analýze a modelování složitých vzorů v datech. Hluboké učení je klíčové pro moderní generativní modely, protože umožňuje efektivní učení a generování realistických obrazů.

Vícevrstvenost v hlubokém učení umožňuje neuronovým sítím rozpoznávat složité vzory postupným zpracováváním dat v několika vrstvách. Když neuronová síť analyzuje obraz konkrétního zvířete, například kočky, vstupní vrstva přijímá obraz a převádí jej na numerická data. Zjednodušeně lze říci že první skrytá vrstva identifikuje základní prvky, jako jsou hrany a jednoduché tvary, například obrysy uší. Druhá skrytá vrstva rozpoznává složitější struktury, jako jsou oči a nos, zatímco třetí skrytá vrstva analyzuje kombinace těchto tvarů, aby identifikovala části těla zvířete, jako je hlava a tělo. Nakonec výstupní vrstva spojuje všechny rozpoznané vzory a určuje, zda je na obraze kočka, čímž poskytuje konečný výsledek. Tímto postupným zpracováváním v několika vrstvách síť dosahuje schopnosti rozpoznávat a klasifikovat složité obrazy [4].

1.5 Generativní umělá inteligence

Generativní umělá inteligence je specifická oblast umělé inteligence zaměřená na tvorbu nového obsahu. Na rozdíl od diskriminačních modelů, které se používají k třídění a rozpoznávání vzorů v datech, generativní modely vytvářejí nové datové vzorky, které jsou podobné těm, na kterých byly trénovány [5].

1.6 Difúzní model

Difuzní model je navržen tak, aby generoval nová data podobná těm, které viděl během tréninku. Tento generativní proces je řízen přidáváním a odstraňováním šumu. V dopředné fázi model přidává šum do obrazu, což postupně přeměňuje obraz na šumový. V reverzní fázi model postupně odstraňuje šum z šumového obrazu, čímž obnovuje původní nebo generuje nový obraz. Tento proces využívá prediktor šumu, který předpovídá množství šumu v každém kroku [5].

1.7 Model latentní difúze

Model latentní difúze vylepšuje tradiční difuzní modely tím, že pracuje v latentním prostoru, což je nižší dimenzionální prostor - složité informace (jako obrázky) jsou převedeny do jednoduššího a nižšího dimenzionálního formátu. Tento formát obsahuje základní charakteristiky původních dat, ale ve zjednodušené podobě. Variační automatický kodér (VAE) komprimuje obraz do latentního prostoru, kde probíhá difúze. Toto výrazně snižuje výpočetní náročnost a zvyšuje kvalitu generovaných obrazů. po dokončení difúzního procesu v latentním prostoru je latentní reprezentace obrazu dekomprimována zpět do původního obrazového prostoru pomocí dekodéru VAE [5].

1.8 Prompt

Prompt je textový popis, který specifikuje požadavky na výsledný obraz. Můžeme také říct, že se jedná o jakýsi typ příkazové řádky pro model AI. Slouží jako vstupní data pro generativní model, který na jeho základě vytváří obraz. Prompt může být jednoduchý, jako například "pes v boudě", nebo složitý a detailní, zahrnující popisy objektů, prostředí a stylu obrazu.

1.9 Datová sada

Datová sada je kolekce dat používaná pro trénování a testování modelů strojového učení. V případě generativních modelů pro obrázky datová sada obvykle obsahuje velké množství obrazů a jejich popisů. Kvalita a velikost datové sady mají zásadní vliv na schopnost modelu generovat kvalitní a realistické obrázky.

2 Stable diffusion a instalátor EasyDiffusion

Stable Diffusion je jedním z nejpokročilejších nástrojů pro generování obrázků pomocí umělé inteligence. Jednou z velkých výhod je, že se jedná o free open source software, tedy jeho použití je zcela zdarma, včetně přístupu ke zdrojovým kódům. Tento systém, jak už víme, využívá difuzní model, který umožňuje transformovat textové popisy do vysoce kvalitních obrázků. Stable Diffusion je navržen tak, aby jej bylo možné provozovat lokálně na relativně běžném počítači uživatele [7].

EasyDiffusion je instalační nástroj a uživatelské rozhraní, které výrazně zjednodušuje instalaci, aktualizace a použití Stable Diffusion. Umožňuje tak uživateli rychle a snadno začít s generováním obrázků, aniž by musel procházet složitými konfiguračními procesy a sháněním správných verzí závislostních komponent, ze kterých je celá AI složena. po instalaci poskytuje EasyDiffusion intuitivní rozhraní, ve kterém mohou uživatelé zadávat textové popisy a okamžitě vidět výsledné obrázky [8].

Díky kombinaci výkonného generování obrázků pomocí Stable Diffusion a jednoduchosti použití EasyDiffusion se tento nástroj stává ideálním řešením pro kohokoli, kdo chce využít sílu umělé inteligence pro tvorbu vizuálního obsahu.

2.1 Výhody lokálního běhu AI oproti běhu v cloudu

2.1.1 Soukromí a bezpečnost dat

Z logiky věci při lokálním provozu AI Stable Diffusion vyplývá, že veškerá data zůstávají na vlastním infrastruktuře (počítači, síti) uživatele. To je obzvláště důležité pro citlivé nebo důvěrné projekty. Uživatelé mají plnou kontrolu nad svými daty, což minimalizuje riziko jejich úniků nebo neoprávněného přístupu k nim.

2.1.2 Nezávislost na internetovém připojení

Používání lokální generativní AI jako je Stable Diffusion eliminuje závislost na internetovém připojení. To znamená, že uživatel může generovat obrázky kdykoliv a kdekoliv, aniž by se musel starat o kvalitu nebo stabilitu internetového připojení.

2.1.3 Výkon a rychlost

Lokální instalace umožňuje využití plného výkonu pořízeného hardwaru uživatele. To může vést k rychlejšímu generování obrázků ve srovnání s některými online službami, které mohou mít omezení v přiděleném výpočetním výkonu, z důvodu nutnosti obsluhy velkého množství uživatelů. Pokud má uživatel výkonný počítač, případně vybavený výkonnou grafickou kartou či dokonce akcelerační kartou pro AI, může dosáhnout značně lepších výsledků.

Další výhodou Stable Diffusion týkající se výkonu je, že není obecně nutný specializovaný výkonný hardware jako podmínka pro jeho použití, protože i na běžném počítači bez použití výkonných akceleračních komponent je možné jeho provoz jen na procesoru, což se projeví na době generování. Řádově se jedná o jednotky minut, o které se proces generování prodlouží, což není pro občasné používání této AI žádná zásadní překážka. Tuto vlastnost mohou ocenit např. právě vysokoškolští studenti, kteří nemají na to, aby si pořídili specializovaný hardware.

Požadavky Stable Diffusion na hardware detailněji rozepisuji v kapitole „Hardwarové požadavky na provoz AI Stable Diffusion“.

2.1.4 Cena

Jak již bylo zmíněno, Stable Diffusion i EasyDiffusion jsou free open source software projekty, tudíž jsou zcela svobodně a zdarma k jakémukoliv použití. Uživatelé nemusí platit za licenci nebo předplatné, což je velkou výhodou oproti některým online službám, které mohou vyžadovat pravidelné poplatky. Tato bezplatnost činí lokální provoz ekonomicky výhodným řešením pro široké spektrum uživatelů, mezi které budou patřit třeba studenti.

Další z výhod ohledně ceny je, že je přístup i ke zdrojovým kódům, za které by například výzkumníci či vývojáři u jiných komerčních projektů museli platit specializované drahé licence.

2.1.5 Možnosti generování

Lokální provoz Stable Diffusion s instalátorem EasyDiffusion poskytuje uživatelům násobně více možností při generování obrázků ve srovnání s online verzemi, a to jak v bezplatné, tak i v placené formě. Lokální řešení umožňuje plný přístup k veškerým

funkcím a parametrům modelu, což zahrnuje detailnější nastavení generování, jemné ladění parametrů, používání rozličných modelů včetně vlastních zahrnující i vlastní datové sady, a neomezené experimentování bez jakýchkoli omezení stanovených online platformami. To umožňuje mnohem větší flexibilitu a kreativitu při práci s modelem, což může být pro mnoho uživatelů zásadní výhodou. Online služby často omezují přístup k pokročilým funkcím nebo požadují za jejich využití dodatečné poplatky, což může být velmi limitující. Lokální provoz tak poskytuje plnou kontrolu a maximální využití všech dostupných nástrojů a možností bez jakéhokoliv omezení, což se týká i množství generovaných obrázků. To je na cloudových nástrojích v bezplatné verzi omezeno počtem kreditů, které uživatel registrací získá, další si musí dokoupit, nebo počkat, pokud se třeba nějaké kredity zdarma přidělují za určitý časový interval.

Dostupné modely ke stažení

Velkou výhodou lokálně provozované AI Stable Diffusion je široká škála dostupných různorodých modelů k volnému použití.

Balík různých funkcí generativní AI v jednom

Další benefitem Stable Diffusion je, že v sobě sdružuje velké množství různých funkcí generativní AI, což u cloudových služeb není běžné, ty jsou obvykle specializovány na konkrétní úlohu. Jedna generuje obrázky, jiná umí upscalling, další upravovat vlastní obrázek a měnit ho na jiný apod. Stable Diffusion obsahuje tyto služby všechny v jednom „balení“ a jsou tak uživateli k dispozici na jednom místě.

Možnost vytváření vlastních modelů a kombinace stávajících

Dalším zvýhodněním Stable Diffusion oproti cloudovým službám je zabudovaný nástroj, umožňující kombinovat různé modely do jiného a využívat tak nově vytvořených modelů nebo také vytvářet modely své vlastní. Taková AI se pak stává mocným nástrojem, pro vytváření děl profesionálních umělců nebo grafiků, které vycházejí z jejich vlastní tvorby a velmi rychle generují nová díla, která jsou věrná stylu autora. Takové funkce cloudové služby obvykle postrádají, protože modely „šité na míru“ individuálně pro konkrétního uživatele jsou u služby, určené masám doslova nežádoucí.

2.2 Nevýhody oproti online prostředí

2.2.1 Požadavky na hardware

Závislost na hardwaru uživatele může být částečně i nevýhodou lokálního provozu. pro profesionální provoz je zapotřebí výkonnějšího hardware a to hlavně pro urychlení generování, protože čekat na vygenerování jednoho obrázku s větším rozlišením řádově větší množství jednotek minut, může být omezující při práci např. profesionálního grafika. je pak samozřejmě otázkou, zda je lepší investovat do cloudové služby, nebo do výbavy vlastního výkonného hardware.

2.2.2 Instalace a údržba

Lokální provoz vyžaduje, aby uživatel provedl instalaci softwaru na svém počítači, případně aby vybral vyhrazený hardware pro konkrétní instalaci a provoz lokální AI. Kromě toho může být potřeba provádět údržby systému ve chvíli, kdy se objeví novější verze, ať už vlastního software AI nebo ovladačů pro hardware akcelerátory ap. Což je nesporně náročnější než prostě zadat adresu cloudové služby do webového prohlížeče.

Naštěstí je proces instalace a aktualizace Stable Diffusion, který sám o sobě není nepřekonatelným problémem, je tento proces velmi zjednodušen instalátorem EasyDiffusion. Jediným zádrhelem se může stát, když musí uživatel vyřešit správný běh za pomoci akcelerace na grafické kartě (GPU) nebo specializované akcelerační kartě. Lokální provoz AI totiž vyžaduje ideálně nejen vysoký výkon, ale je nutné poměrně velké množství RAM, nejen v podobě operační paměti, ale také vyhrazené vlastní RAM pro GPU nebo akcelerační kartu. Pokud uživateli stačí běh na CPU nebo má „štěstí“, že mu běží rovnou akcelerace na GPU, které již v počítači má je pak instalace a údržba Stable Diffusion v podstatě jen spuštění EasyDiffusion, který se při prvním spuštění postará o instalaci a při každém dalším o aktualizace.

2.3 Parametry generování obrázku

Prostředí EasyDiffusion poskytuje uživatelům širokou škálu parametrů, které mohou být užity pro optimalizaci výsledků při generování obrázků. Tyto parametry umožňují uživatelům kontrolovat různé aspekty procesu generování a přizpůsobit výstupy jejich specifickým potřebám. Níže jsou uvedeny hlavní parametry, které mohou být nastaveny.

2.3.1 Prompt (výzva)

Textový popis je základním vstupem pro generování obrázku. Uživatelé zadávají text, který popisuje, co má být na obrázku vygenerováno. Stavba a struktura promptu jsou klíčové pro výsledný obrázek. Následující obrázky znázorňují jak pouhé čárky oddělující části promptu mohou ovlivnit výsledek [9].

Prompt: closeup photo of brown cat in the mug which is in the kitchen lit by daylight

Výsledný obrázek:



Obrázek 1: Vygenerovaný obrázkem promptem "closeup photo of brown cat in the mug which is in the kitchen lit by daylight"

Prompt: brown cat in the mug, closeup photo, kitchen background, dayligth

Výsledný obrázek:



Obrázek 2: Vygenerovaný obrázek po zadání promptu "brown cat in the mug, closeup photo, kitchen background, dayligth"

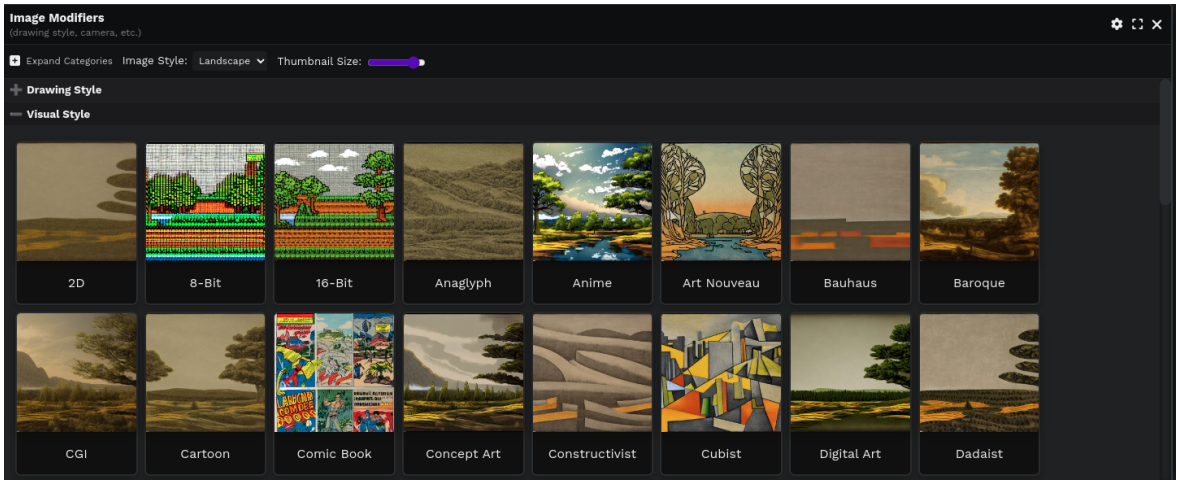
Varianty promptu jsou různé verze stejného základního promptu, které mohou generovat odlišné výsledky. Zadáváme je do složených závorek [9].

Váhy promptu určují důležitost jednotlivých slov nebo frází v promptu. Čím vyšší váhu přiřadíme, tím větší vliv bude mít na generovaný obrázek. Váhy nám umožňují jemně doladit výsledky a zvýraznit určité aspekty obrazu. Váhy definujeme znaménky + a – nebo čísly [9].

2.3.2 Image Modifiers (modifikátory obrázku)

Modifikátory umožňují upravit výsledný styl obrázku přidáním specifických atributů, jako jsou styly vykreslení, barvy, techniky a mnoho dalších [10].

Dostupná nabídka modifikátorů může také posloužit jako inspirace pro sestavení promptu.



Obrázek 3: Ukázka modifikátorů obrázku v EasyDiffusion

2.3.3 Seed (semínko)

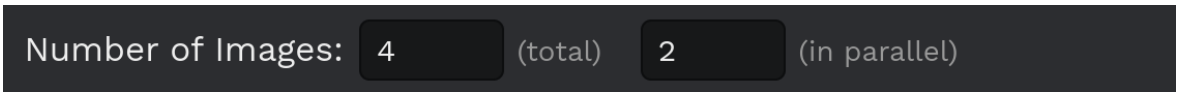
Semínko je náhodné číslo, které inicializuje generátor. Použití stejného čísla s identickými parametry povede ke stejnému výslednému obrázku, což je užitečné pro replikaci a jemné ladění výsledků.



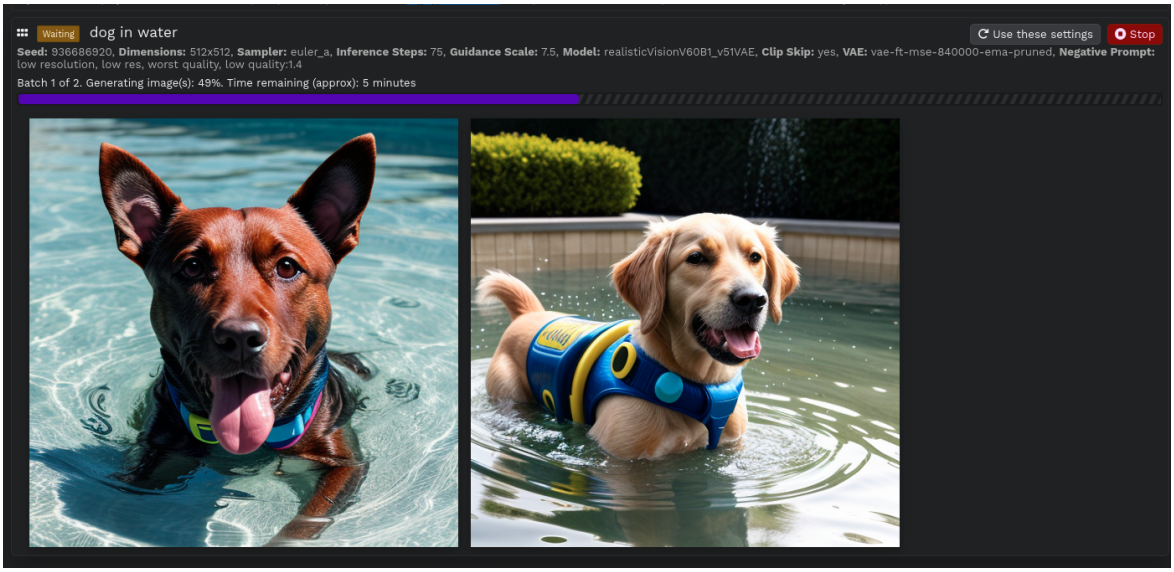
Obrázek 4: Ukázka parametru Seed v EasyDiffusion

2.3.4 Number of Images (počet obrázků)

Uživatelé mohou nastavit, kolik obrázků má být vygenerováno a v kolika dávkách. Např. vyplníme, že chceme vygenerovat celkem 4 obrázky ve 2 dávkách. Budou se tedy vygenerovat 2 obrázky najednou. To je užitečné pro porovnávání různých variant a výběr nejlepšího výsledku.



Obrázek 5: Ukázka parametru Number of Images v EasyDiffusion



Obrázek 6: Ukázka generování 2 obrázku najednou v EasyDiffusion

2.3.5 Model

EasyDiffusion umožňuje uživatelům vybrat specifický model nebo verzi modelu, který bude použit pro generování obrázku. Různé modely mají různé schopnosti a styly, což poskytuje další úroveň přizpůsobení.

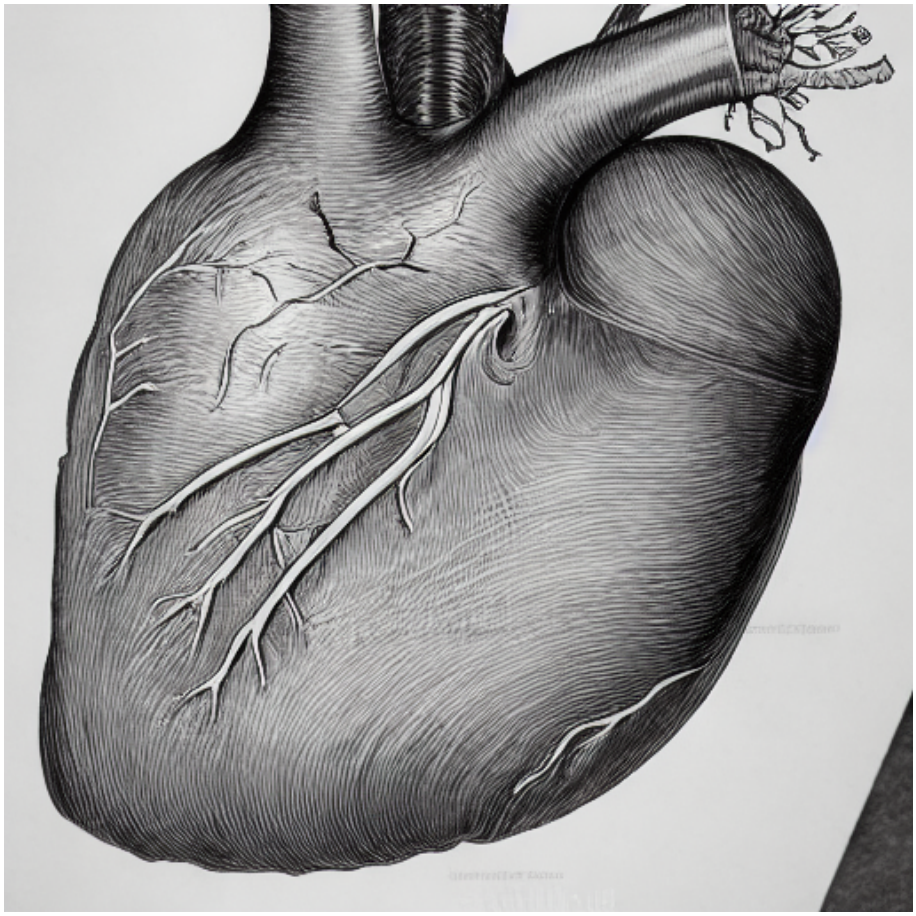
Při nainstalování prostředí EasyDiffusion je zároveň nainstalován i výchozí model Stable Diffusion v1.5 (sd-v1-5). Při generování však s velkou pravděpodobností bude třeba rozšířit množství dostupných modelů.

Stable Diffusion v1.5

Stable Diffusion v1.5 je pokročilý model generování obrazů zaměřený na tvorbu převážně realistických obrázků. Tento model je ideální pro širokou škálu požadavků a je oblíbený právě pro svou flexibilitu a schopnost přizpůsobit se různým stylům. To má však někdy za následek větší pracnost, při vytváření promptů a ladění jednotlivých parametrů.

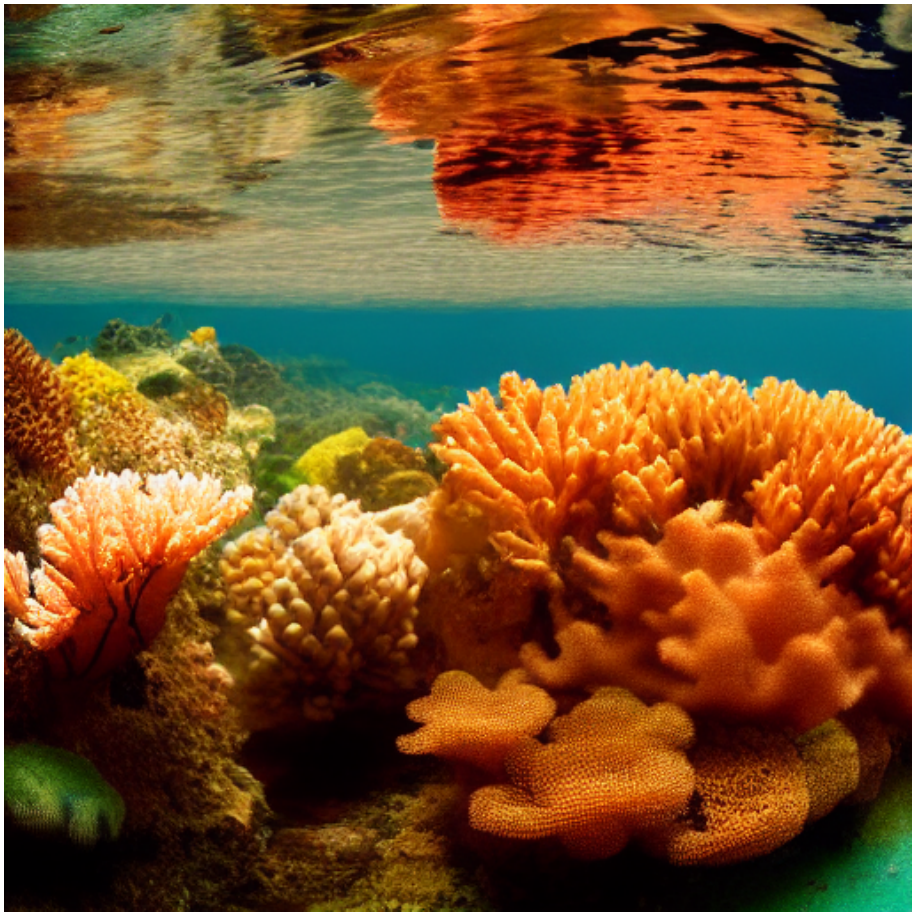
V kombinaci s tímto modelem se mi osvědčily samplery Euler Ancestral a DDIM. Zároveň umí pracovat spíše v menším rozlišení, konkrétně 512×512, při vyšším počtu iterací je schopen i rozlišení 768×768. Jinak mu stačí menší počet kroků (podle náročnosti promptu se počet iterací při rozlišení 512×512 pohyboval od 15 do 30).

- realistic illustration of heart organ, detailed sketch, human anatomy
  low quality, low resolution, low res



Obrázek 7: Vygenerovaný obrázek po zadání promptu "realistic illustration of heart organ, detailed sketch, human anatomy"

- photography of underwater scenery++, glowing corals, fishs, Colorful low quality, low resolution, low res



Obrázek 8: Vygenerovaný obrázek po zadání promptu "photography of underwater scenery++, glowing corals, fishs, Colorful"

- high resolution photography interior design, a cozy scandinavian living room, wooden floor, many details, studio light



Obrázek 9: Vygenerovaný obrázek po zadání promptu "high resolution photography interior design, a cozy scandinavian living room, wooden floor, many details, studio light"

Realistic Vision

Realistic Vision je model navržený pro generování hyper-realistických obrazů. Tento model je jednou z nejlepších voleb, pokud je cílem vytvořit detailní a věrné fotografie lidí, krajin ap., které vyžadují vysokou úroveň realismu. Realistic Vision totiž vyniká v práci s jemnými detaily, jako jsou textury pleti, vlasy a osvětlení, což umožňuje dosažení výsledků, které jsou téměř k nerozeznání od skutečných fotografií. je proto ideální pro projekty, kde je kladen důraz na věrohodnost výstupu.

Tento model se mi nejvíce osvědčil v kombinaci se samplerem DPM++ 2m SDE (Karras).

- a woman with dark hair, unforgettable, high quality



Obrázek 10: Vygenerovaný obrázek po zadání promptu "a woman with dark hair, unforgettable, high quality"

- a small balcone with the Eiffle Tower view, morning sun, amazing breakfast, high qualit

  low quality, low resolution, low res



Obrázek 11: Vygenerovaný obrázek po zadání promptu "a small balcone with the Eiffle Tower view, morning sun, amazing breakfast, high qualit"

- photography of big (crocodile)+ in water, high quality low quality, low resolution, low res, mutated hands and fingers, extra fingers, bad anatomy



Obrázek 12: Vygenerovaný obrázek po zadání promptu "photography of big (crocodile)+ in water, high quality"

DreamShaper

DreamShaper je model navržený pro tvorbu snových, fantastických a surrealistických obrazů. Umožňuje vytvářet jedinečné a imaginativní scény s abstraktními tvary, zářivými barvami a fantastickými motivy. Tento model podporuje kreativitu a experimentování. Pro studenty umění, literatury nebo psychologie může DreamShaper nabídnout cenný nástroj pro vizualizaci a tvorbu uměleckých projektů, které vyžadují kreativní a neotřelý přístup.

- oil painting of a future city of (flying islands)+, fantasy sci - fi city, fantasy style, highly detailed, magnificent

  low quality, low resolution, low res



Obrázek 13: Vygenerovaný obrázek po zadání promptu "oil painting of a future city of (flying islands)+, fantasy sci - fi city, fantasy style, highly detailed, magnificent"

- a 3D model of one cute raspberry fruit, big smile, sharp teath, clean background low quality, low resolution, low res

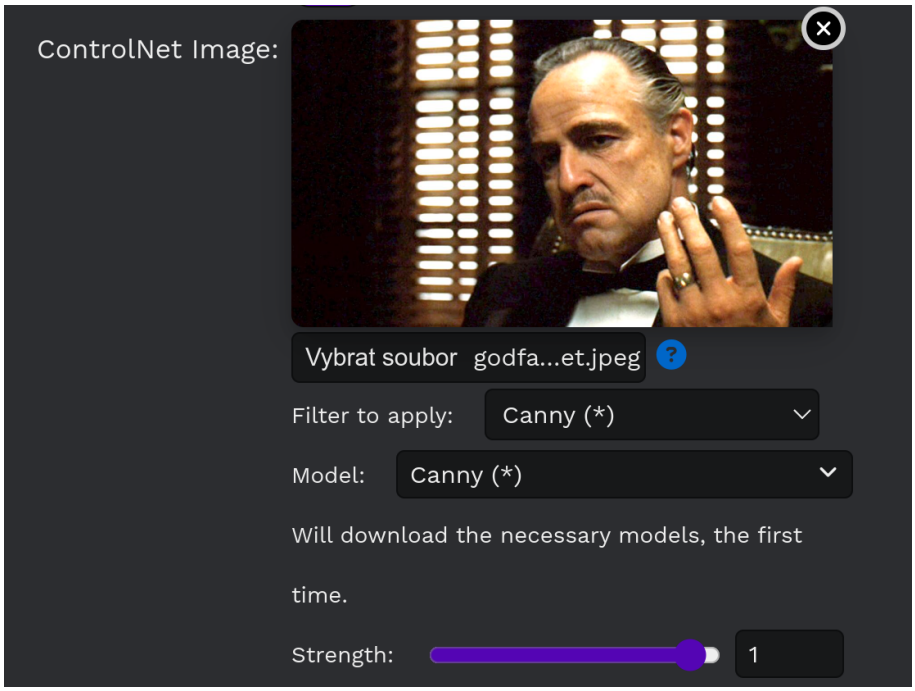


Obrázek 14: Vygenerovaný obrázek po zadání promptu "a 3D model of one cute raspberry fruit, big smile, sharp teath, clean background"

2.3.6 ControlNet

Tento parametr umožňuje použít kontrolní obrázek k vedení AI při generování nových obrázků. Tento parametr pomáhá AI přesněji následovat vybraný styl, tělesnou pózu nebo kompozici z kontrolního obrázku. Na rozdíl od image-to-image transformace, která aplikuje úpravy na původní obrázek, ControlNet dokáže extrahovat smysluplné informace z kontrolního obrázku a použít je k vytvoření zcela nových obrázků v podobném stylu [11].

Pro ukázku tohoto parametru jsem zvolila jako výchozí obrázek, ze kterého bude generátor čerpat kompozici, styl a tělesnou pózu, známý snímek z filmu Kmotr. Prompt, který jím bude ovlivněn zní Johnny Depp. Filtr spolu s modelem ControlNet parametru pro tyto účely nastavíme na Canny.



Obrázek 15: Ukázka nastavení funkce ControlNet Image v EasyDiffusion

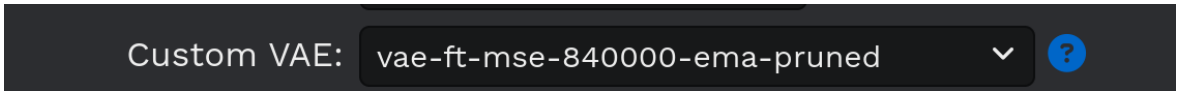
Výsledný obrázek přebral kompozici a přesné rozlišení výchozí fotografie.



Obrázek 16: Vygenerovaný obrázek prostřednictvím funkce ControlNet - JohnnyDepp

2.3.7 VAE (Variational Autoencoder)

VAE zlepšuje kvalitu generovaných obrázků. je trénováno na různé aspekty obrazu. Výchozí VAE model v uživatelském rozhraní EasyDiffusion (vae-ft-mse-840000-ema-pruned) specificky vylepšuje detaily očí na generovaných snímcích [12].

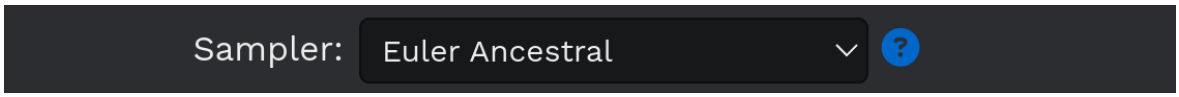


Obrázek 17: Ukázka vyplnění parametru VAE v EasyDiffusion

2.3.8 Sampler (vzorkovač)

EasyDiffusion nabízí různé "samplery", které mohou produkovat odlišné výsledné obrázky. Každý sampler má vlastní unikátní metodu, jak převést vstupní data na konečný obraz [13].

Je dobré vyzkoušet různé samplery, abyste zjistili, který z nich nejlépe odpovídá vašim potřebám. V tomto případě nelze přesně predikovat a určit, který sampler používat pro jaký styl obrázku, jako je to třeba u modelu. Různé obrázky mohou vypadat lépe s různými samplery. V kapitole modelů jsem se však snažila vybrat spojení modelů a samplerů, které spolu dobře fungují.



Obrázek 18: Ukázka vyplnění parametru Sampler v EasyDiffusion

2.3.9 Image Size (velikost obrázku)

Uživatelé mohou nastavit šířku a výšku výsledného obrázku. Vyšší rozlišení produkuje detailnější obrázky, ale také vyžaduje více výpočetního výkonu, času a použití modelu, který s daným rozlišením umí pracovat. Běžné nastavení zahrnuje šířku a výšku v pixelech (např. 512×512, 1024×1024).



Obrázek 19: Ukázka vyplnění parametru Image Size v EasyDiffusion

2.3.10 Inference Steps (počet kroků)

Počet kroků určuje, v kolika iteracích proběhne vygenerování obrázku. Více iterací obvykle vede k lepším výsledkům, ale také zvyšuje čas potřebný k vygenerování obrázku.



Obrázek 20: Ukázka vyplnění parametru Inference Steps v EasyDiffusion

2.3.11 Guidance Scale (naváděcí škála)

Tento parametr určuje, jak silně model sleduje zadaný textový popis. Vyšší hodnoty vedou k obrázkům, které jsou více v souladu s popisem, ale mohou také omezit kreativitu modelu. Hodnoty se obvykle pohybují mezi 5 a 20. Stable Diffusion nabízí stupnici od 1,1 do 50.



Obrázek 21: Ukázka vyplnění parametru Guidance Scale v EasyDiffusion

2.3.12 Seamless Tiling (bezešvé dlaždice)

Jedná se o techniku, která umožňuje vytváření vzorů, které se mohou opakovat bez viditelných spojů nebo okrajů, což vytváří hladký a jednotný vzhled. Tato může být žádané při tvorbě opakujících se textur pro herní prostředí, pozadí plakátů, tapet ap. [14]

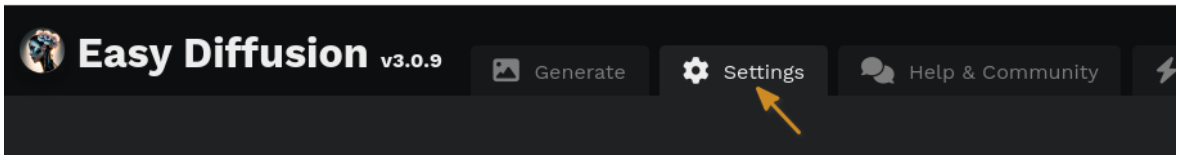
Vygenerovanou dlaždici a její kopie lze vedle sebe umístit tak, aby vytvořily plynulý, opakující se vzor bez viditelných přechodů. je možné vybírat mezi horizontálním, vertikálním a přechodem v obou směrech.

2.3.13 Output Format (výstupní formát)

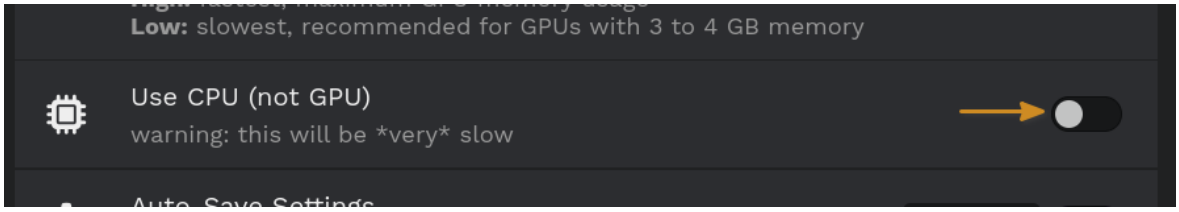
Možnost volby formátu výstupních obrázků, jako jsou PNG, JPEG nebo WEBP včetně možnosti zapnutí bezztrátové komprese.

2.3.14 Přepnutí akcelerovaného běhu na běh na procesoru

Ve webovém rozhraní EasyDiffusion jde v nastavení konfigurace přepnout generování obrázků z akcelerovaného běhu (automatický výběr nejrychlejšího způsobu) na běh čistě na procesoru.



Obrázek 22: Navigace na možnost Settings v EasyDiffusion



Obrázek 23: Ukázka možnosti přepnutí akcelerovaného běhu na běh na procesoru

2.4 Text to image a image to image metody generování obrázku

Text to Image (T2I) je metoda generování obrázků, která využívá textové popisy jako vstup k vytvoření vizuální reprezentace. Pomocí modelu Stable Diffusion mohou uživatelé zadávat podrobný textový popis, který model následně přemění na obrázek. Výše jsme si tento proces a jeho kroky popsali.

Image to Image (I2I) je metoda generování obrázků, která využívá existující obrázek jako vstup a transformuje jej podle zadaného textového popisu nebo dalších parametrů. S modely Stable Diffusion mohou uživatelé nahrát počáteční obrázek a následně jej upravovat nebo stylizovat na základě specifických požadavků.

V rámci této práce je rozebírána především metoda T2I.

3 Hardwarové požadavky na provoz AI Stable Diffusion

3.1 Provoz softwarově na procesoru

Stačí běžné PC s procesorem architektury x86-64 (neboli amd64) nebo Mac s ARM (M1 / M2 / M3) architekturou. Obecně platí, čím výkonnější CPU, čím více jader a čím více vláken na jádro, tím lépe, zpracování bude rychlejší. u pomalejšího stroje si uživatel adekvátně delší dobu počká, než se mu obrázek vygeneruje. Důležité je mít dostatek RAM, aby se do ní vešel celý model a také veškerá data, při generování obrázku. Proto je minimum pro rozumný provoz 16 GiB RAM nebo více. i zde platí, čím více, tím lépe, je pak možné používat větší modely, vyšší rozlišení generovaného obrázku, složitější výzvy ap.

3.2 Akcelerace pomocí GPU

Akcelerovat běh AI Stable Diffusion na grafické kartě lze v principu na čemkoli, co podporuje TensorFlow (bezplatná softwarová knihovna s otevřeným zdrojovým kódem pro strojové učení a umělou inteligenci vyvinutá týmem Google Brain) a PyTorch (knihovna strojového učení založená na knihovně Torch, používaná pro aplikace, jako je počítačové vidění a zpracování přirozeného jazyka, původně vyvinutá Meta AI a nyní součástí zastřešující Linux Foundation). Prakticky to znamená použití grafických karet NVIDIA (přes API CUDA), AMD (přes API ROCm) nebo GPU z architektury M1 / M2 / M3 (přes API Metal) firmy Apple. i zde podobně jako u běhu na CPU je důležité mít pro grafický čip k dispozici dostatečně velkou VRAM. Přestože je minimálním doporučením přímo od autorů EasyDiffusion grafická karta se 4 GiB VRAM, je toto pro praktické použití málo. Ideální je alespoň 8 GiB VRAM pro generování obrázků v rozlišení do 512×512 pixelů se standardními modely. pro větší a složitější modely (např. SD XL) a větší rozlišení obrázků, typicky 768×768 pixelů nebo 1024×1024 pixelů už nastává potřeba 12 GiB VRAM nebo ideálně 16 GiB. Výhodou jsou zde např. procesory AMD Ryzen s dostatečně výkonnými integrovanými GPU, kterým je možné v BIOS/UEFI Setupu přidělit až 16 GiB VRAM. Takto nakonfigurovaný PC pak zvládá i úlohy, které není možné zpracovat ani na velmi výkonné (a drahé) grafické kartě, z důvodu jejího menšího objemu VRAM integrovaného na kartě. Dá se říct, že požadavek na dostatek vyhrazené RAM je stěžejní a je dobré na něj dávat důraz při výběru vhodné grafické karty pro urychlení běhu AI.

3.3 Akcelerace pomocí specializovaných čipů

Akcelerovat běh AI Stable Diffusion je možné také pomocí specializovaných čipů. i zde jak již bylo zmíněno výše je v principu podporováno vše, co podporuje akcelerovaný běh knihovny TensorFlow a PyTorch. Dobrým příkladem takového specializovaného HW jsou např. akcelerační karty Xilinx Alveo firmy AMD nebo AI accelerator application-specific integrated circuit (ASIC) TPU (Tensor Processing Unit) koprocesory firmy Google, vyvinuté přímo pro akcelerovaný běh knihovny TensorFlow a PyTorch. i zde platí obdobné požadavky na velikost vyhrazené RAM pro akcelerační čip. Ideální jsou proto akcelerační karty se 16 GiB RAM a více. Různé akcelerátory pro strojové učení, které jsou běžně k sehnání do PCIe sběrnice nebo na USB 3 nejsou pro tento typ úloh – jakým je generování obrázků – vhodné, protože mají zpravidla nedostatek vlastní RAM, jsou určeny pro jiné typy úloh založených na strojovém a hlubokém učení.

3.4 Proč je běh umělé inteligence na GPU nebo AI akcelerátorech rychlejší

Z výše uvedených informací v této sekci, se nabízí otázka, proč je běh umělé inteligence na grafických nebo na specializovaných čipech o tolik rychlejší, že se používá pro akceleraci běhu AI?

Rychlá velmi zjednodušená odpověď je, že umělé neuronové sítě, na kterých jsou postavené dnešní algoritmy umělé inteligence, založené na strojovém a hlubokém učení, jsou typickým příkladem pro paralelní zpracování jednotlivých podúloh. Co to ale znamená vzhledem k procesorů, grafickým kartám a specializovaným čipům? Dnešní běžné procesory do osobních počítačů mají 2 až 16 jader, často s možností běhu 2 vláken na každé jádro. Takový procesor se pak chová jakoby by vybaven dvojnásobným počtem výpočetních jader. Dá se proto říct, že procesory v dnešním osobních počítačích mají až několik desítek virtuálních výpočetních jader. Grafické procesory mají ale i stovky výpočetních jader a navíc specializovaných pro běh úloh, které sedí právě algoritmům pro umělou inteligenci. Obdobně to je u AI akcelerátorů, ať už se jedná o specializované ASIC čipy, které jsou navrženy explicitně pro běh těchto konkrétních úloh nebo velká programovatelná hradlová pole, která po příslušném naprogramování vlastně vytvoří hromady specializovaných

výpočetních jader pro běh neuronových sítí. GPU a specializované čipy tak těží z několika výhod oproti univerzálním počítačovým procesorům:

- větší množství výpočetních jader
- specializace výpočetních jader na konkrétní úlohy
- rychlejší a specializované dedikované RAM (vyšší frekvence a typ paměti GDDR a HBM)
- rychlejší spojení s dedikovanou RAM, prostřednictvím širší datové sběrnice (až 512 bitů u GDDR a až 4096 bitů u HBM pamětí oproti 64 bitové datové paměťové sběrnici, kterou používají procesory ve spojení s DDR paměťmi)

3.5 Porovnání řešení vzhledem k výkonu

Pro ukázku vlivu výkonu hardware na rychlost generování obrázků jsem zvolila 3 běžné PC sestavy, u kterých jsem testovala nenáročný prompt s výchozím modelem. Generování obrázku jsem spustila na procesoru a také akcelerovaně na grafické kartě (pokud to sestava umožňovala). V testu mi nešlo o porovnání nejvýkonnějších dostupných řešení s drahými dedikovanými grafický kartami, které vygenerují i více než 1 obrázek za sekundu, ale o vyzkoušení běžných osobních počítačů, vlivu zrychlení generování obrázků běžícího na grafickém čipu místo na procesoru a vlivu rychlosti pamětí. u běžných sestav rozdíly lépe vyniknou a dají lepší představu o rozdílech ve výkonu, než by to bylo u velmi výkonných řešení, která se u generování obrázku liší řádově jen v desetinách sekund, maximálně v sekundách.

3.5.1 Testovací PC sestava 1

Jako první sestavu jsem zvolila dnes již zastaralý hardware s procesorem Intel Core 4. generace, který i v době kdy byl nový, byl určen pro běžnou kancelářskou práci, nikoli do výkonné pracovní stanice. PC mi nyní slouží jako serverový a multimediální server v domácí síti. Vyzkoušet, jak si takový zastaralý stroj poradí s během moderní generativní umělé inteligence se nabízelo jako vhodný srovnávací základ. Test rychlosti generování na této sestavě byl proveden pouze na procesoru, protože není vybaven pro akcelerovaný běh např. na grafické kartě.

- CPU: Intel Core i3-4160 @ 3,60GHz (4C/4T)
- RAM: 24GiB DDR3 2,133 GHz (2× 8 GiB + 2× 4 GiB Dual Channel)

3.5.2 Testovací sestava 2

Další sestava, kterou jsem do testu zahrnula je poměrně moderní stroj a procesorem AMD Ryzen 5. generace , vybavený i integrovaným grafickým jádrem, které je i přes integraci v procesoru a z toho vyplývající omezení, dostatečně výkonné na akceleraci běhu AI. Skvělou vlastností je, že v BIOS/UEFI setup stroje je možné vyhradit pro grafický čip až 16 GiB VRAM, která je GPU přidělena z běžné RAM počítače.

- CPU + GPU: AMD Ryzen 7 5700G with Radeon Graphics (8C/16T)
- RAM: 32 GiB DDR4 2,667 GHz (4× 8 GiB Dual Channel) – 16 GiB konfigurováno pro VRAM
- Chipset: B450

3.5.3 Testovací sestava 3

Poslední sestava, kterou jsou zvolila pro test, je téměř totožná jako druhá, s malými rozdíly. Procesor s integrovanou grafickou kartou je totožný, ale je vložen do základní desky, která má o generaci novější čipovou sadu, díky které je možné v PC provozovat paměti s vyšší frekvencí než předchozí sestavě. Zajímal mě vliv rychlejších pamětí na rychlost generování obrázku ať již na procesoru, tak na GPU, která používá stejnou RAM.

- CPU + GPU: AMD Ryzen 7 5700G with Radeon Graphics (8C/16T)
- RAM: 64 GiB DDR4 3,6 GHz (2× 32 GiB Dual Channel) – 16 GiB konfigurováno pro VRAM
- Chipset: B550

PC sestava	CPU		GPU	
	čas na iteraci	**čas celkem**	**čas na iteraci**	**čas celkem**
1.	13,55 s / it	6:00 min	-	-
2.	8,44 s / it	3:49 min	2,46 s / it	1:24 min
3.	7,66 s / it	3:21 min	2,09 s / it	1:06 min

Tabulka 1: Záznam měření času generování obrázku na různých PC sestavách

Z výsledku měření generování vyplývají zajímavé výsledky.

První je příjemně překvapující zjištění, že i morálně zastaralý stroj může nabídnout možnost běhu generativního AI, aniž by to bylo pro občasné použití nějak výrazně omezující. Vygenerování obrázku z jednoduchého promptu trvá jen několik minut, s během AI nebyly žádné problémy, 24 GiB RAM je zjevně dostačující pro tento typ úloh. Pustit si proto umělou inteligenci dnes může prakticky každý.

Dalším zajímavým zjištěním je, že běh na výrazně novějším procesoru nevede k předpokládanému zrychlení AI, přestože nový CPU má 2× více jader a je schopen zpracovat dvojnásobný počet vláken na každém jádře oproti starému procesoru. Přičítám to tomu, že při běhu AI softwarově na procesoru nejsou naplno využita všechna procesorová (i virtuální) jádra. Běh na procesoru zde zjevně není moc dobře optimalizován a zvýšení počtu jader nemá výrazný vliv. z monitoringu zatížení procesoru vyplývá fakt, že téměř s plným zatížením běží pouze 2 procesorová jádra (na novém i starém stroji), u ostatních je zatížení výrazně nižší a není trvalé, protože hodně kolísá – nejspíše jen dávkově vykonávají nějakou podružnou práci. Při běhu na CPU je tak vidět pouze vliv novější architektury, které má vyšší počet zpracovaných instrukcí na takt a vyšší frekvence. je tak zde velká rezerva pro další optimalizaci běhu na nových moderních procesorech s větším množstvím procesorových jader a možností zpracování více vláken na každém jádře.

Zjištění jak moc se zrychlí generování obrázku, když je úloha prováděna na grafickém čipu je rovněž zajímavé. i relativně slabý grafický čip, integrovaný v pouzdře procesoru ukazuje

jaký pozitivní vliv na běh AI tento způsob provozu má. Téměř šestinásobné zrychlení nejrychlejší testované sestavy oproti nejpomalejší je moc pěkným výkonnostním bonusem.

Neméně zajímavý je také fakt ohledně vlivu rychlejších pamětí u jinak identických hardwarových sestav. Vyplývá z toho jasná informace, že u běhu takové AI nezáleží čistě na výpočetním výkonu, ale důležitá je také výkon paměťová části stroje, protože zjevně dochází k mnoha operacím, kdy velké množství dat prochází tam a zpět mezi operační pamětí a procesorem. Koneckonců toto vysvětluje i používání pamětí HBM2 v nejvýkonnějších dedikovaných grafických kartách a specializovaných AI akcelerátorech, vliv co nejrychlejších pamětí je zde neodiskutovatelný.

Vyvstává tak otázka jestli pomalý běh na CPU, kdy jsou výpočetní jádra nevyužita, je vůbec řešitelný případnými softwarovými optimalizacemi. Úzkým hrdlem[1] totiž nemusí být hrubý výpočetní výkon, ale relativně malá šířka paměťové sběrnice běžných operačních pamětí, se kterými navíc univerzální počítačový procesor pracuje jiným způsobem, než grafický čip či specializovaný čip pro akceleraci běhu umělé inteligence. Procesor má fakticky samostatná procesorová jádra, která si na paměťové sběrnici vlastně konkurují a navíc zpracovávají velmi malé kousky dat (max několikabajtové sekvence, zvané slova). Grafický čip oproti tomu pracuje s pamětí jako jeden procesor, nedochází tak ke konkurenci jednotlivých výpočetních jader a navíc přistupuje k datům v paměti po výrazně větších blocích naráz, dochází tak k lepšímu využití paměťové komunikační sběrnice a práci s pamětí. Procesor (nebo jeho jednotlivá jádra) musí na data nebo na možnost dostat se k obsazené paměťové sběrnici čekat až se uvolní. Dochází tak ke zpomalení běhu a nevyužití výpočetního výkonu, který je k dispozici. je to daň za univerzálnost, kterou počítačový procesor disponuje.

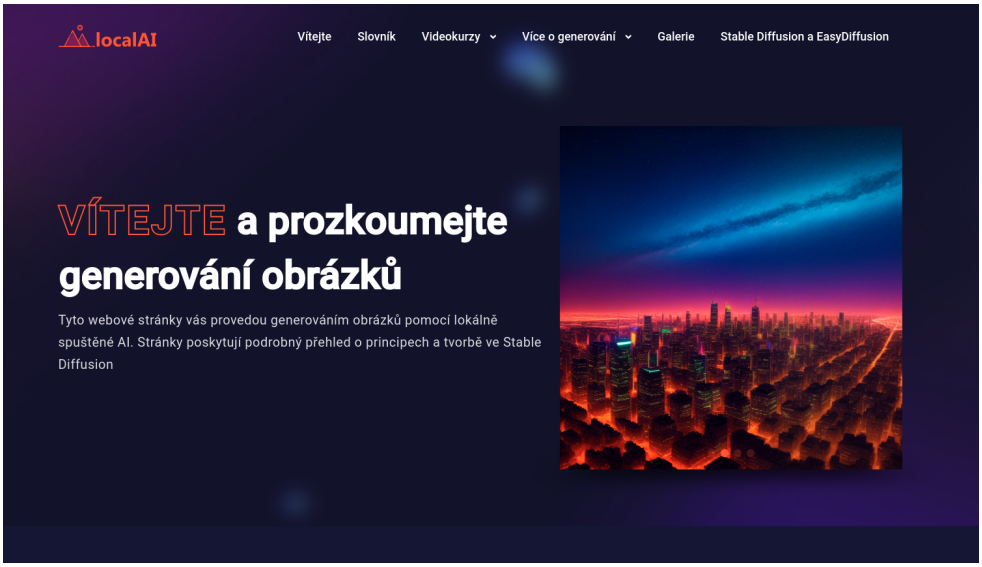
1 Bottleneck, česky úzké hrdlo, známý též efekt hrdla lahve, se ve výpočetní technice označuje takové místo, které je zdrojem zpomalení nebo zahlcení, kdy ostatní části počítače musejí na vybavení operací v tomto úzkém místě čekat, než mohou pokračovat v práci a dochází tak k neefektivnímu využití výpočetního výkonu, protože v tomto zúžení je nejnižší propustnost v rámci celého systému. Z tohoto jevu vznikla parafrázovaná poučka „počítač je tak rychlý, jak je rychlá jeho nejpomalejší část“.

Část II – praktická část

4 Úvod praktické části

V praktické části práce se zaměřuji na tvorbu webových stránek sloužících jako multimediální průvodce generováním obrázků prostřednictvím lokálně provozovaného modelu umělé inteligence. V praktické části shrnuji proces tvorby těchto webových stránek, detailně popisuji vytváření a obsah videoprůvodců, věnuji se také struktuře a obsahu webu samotného.

Odkaz na webové stránky: https://localai.loocajda.name



Obrázek 24: Úvodní stránka webových stránek

4.1 Postup práce

4.1.1 Výběr tématu

Při výběru tématu mé bakalářské práce jsem věděla, že se chci věnovat AI, vzhledem k rychlému vývoji v této oblasti. Baví mě práce s grafikou, takže jsem to propojila a zaměřila se na generování grafiky právě prostřednictvím umělé inteligence. V době výběru tématu bylo velice populární používat k tvorbě obrázků Midjourney. Já jsem si však pro svou práci vybrala Stable Diffusion kvůli jeho flexibilitě a otevřenému zdrojovému kódu.

4.1.2 Vyhledání a studium zdrojů

V této fázi jsem se zaměřila na vyhledání odborné literatury a online zdrojů týkajících se generování obrázků, zejména modelu Stable Diffusion, na kterém je EasyDiffusion založen. Dále jsem analyzovala existující průvodce a tutoriály, které se zabývají praktickými aspekty práce s tímto modelem.

4.1.3 Analýza principu generování obrázků

Abych mohla vytvořit kvalitní průvodce, bylo nezbytné porozumět fungování modelu EasyDiffusion. Zaměřila jsem se především na princip latentní difúze, která je základem pro generování obrázků. Analyzovala jsem architekturu modelu, jeho parametry a možnosti konfigurace. Dále jsem zkoumala, jakým způsobem je model trénován a jakým způsobem generuje nové obrázky.

4.1.4 Položení teoretických základů

Na základě získaných informací jsem sestavila teoretický základ pro svou práci. Tento základ zahrnuje vysvětlení klíčových pojmů, jako je generativní model, latentní prostor, difúzní proces a neuronové sítě. Dále jsem popsala principy fungování modelu Stable Diffusion a jeho implementaci v EasyDiffusion. Teoretický základ sloužil jako pilíř pro další části práce.

4.1.5 Sestavení scénářů a kompletace videonávodů

Vytvořila jsem několik scénářů, které postupně uvádějí do problematiky generování obrázků pomocí EasyDiffusion. Obsah videí jsem skládala tak, aby byl konzumentem co nejlépe pochopen, aby nebyl zahlcen velkým množstvím informací. Mým záměrem bylo také vytvořit úměrně dlouhá videa. Podle scénářů jsem následně zkompletovala videa, ve kterých jsem krok za krokem demonstrovala jednotlivé postupy.

4.1.6 Navržení struktury webu a jeho realizace

Nejen na základě teoretických poznatků a vytvořených videonávodů jsem navrhla strukturu webových stránek. Stránky jsou rozděleny do několika sekcí, které obsahují teoretický základ, praktické návody, galerii generovaných obrázků a další doplňující materiály. Při tvorbě webových stránek jsem kladla důraz na uživatelskou přívětivost a responzivní design, aby byl web přístupný na různých zařízeních.

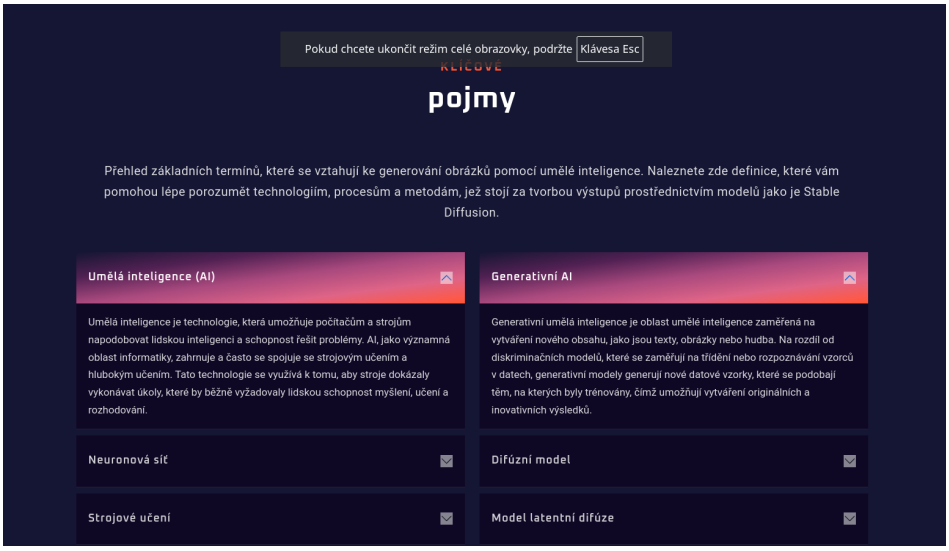
4.2 Hlavní cíle a funkce webových stránek



Obrázek 25: Rozcestník webových stránek

4.2.1 Seznámení s pojmy

Jednu ze stránek webu jsem pojala jako slovník. Uživatel má tedy k dispozici definice pojmů, jež jsou používány na celém webu a jejich znalost je důležitá pro pochopení dalšího obsahu webu, hlavně jednotlivých videonávodů. Při tvorbě slovníku jsem se zaměřila na přesnost, určitou míru stručnosti (aniž by byl potlačen význam), srozumitelnost a také relevanci, aby nedocházelo k zahlcení nadbytečnými informacemi. Jedním z cílů celého webu je totiž přednést téma generování dobře pochopitelnou formou.



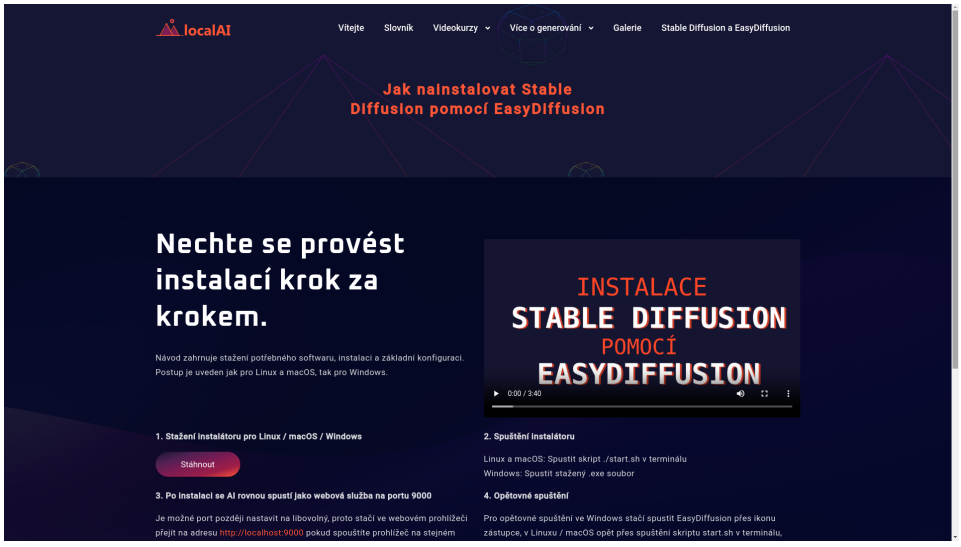
Obrázek 26: Stránka webu se slovníkem

4.2.2 Videoprůvodci

Jednotlivá videa uživatele provedou celým procesem od instalace softwaru až po pokročilé techniky generování obrázků. Uživatel tak získá potřebné informace ke zprovoznění Stable Diffusion, k čemuž použije instalátor EasyDiffusion a dozví se jak funguje celý proces generování. Jeden z videonávodů je věnován tvorbě promptů, což je jedna z hlavních schopností potřebných k dosažení kýžených výsledků při generování. Další schopnost je orientace v modelech, které AI používá – proto jeden z videoprůvodců mapuje a porovnává modely pro EasyDiffusion a předkládá názorné ukázky generování.



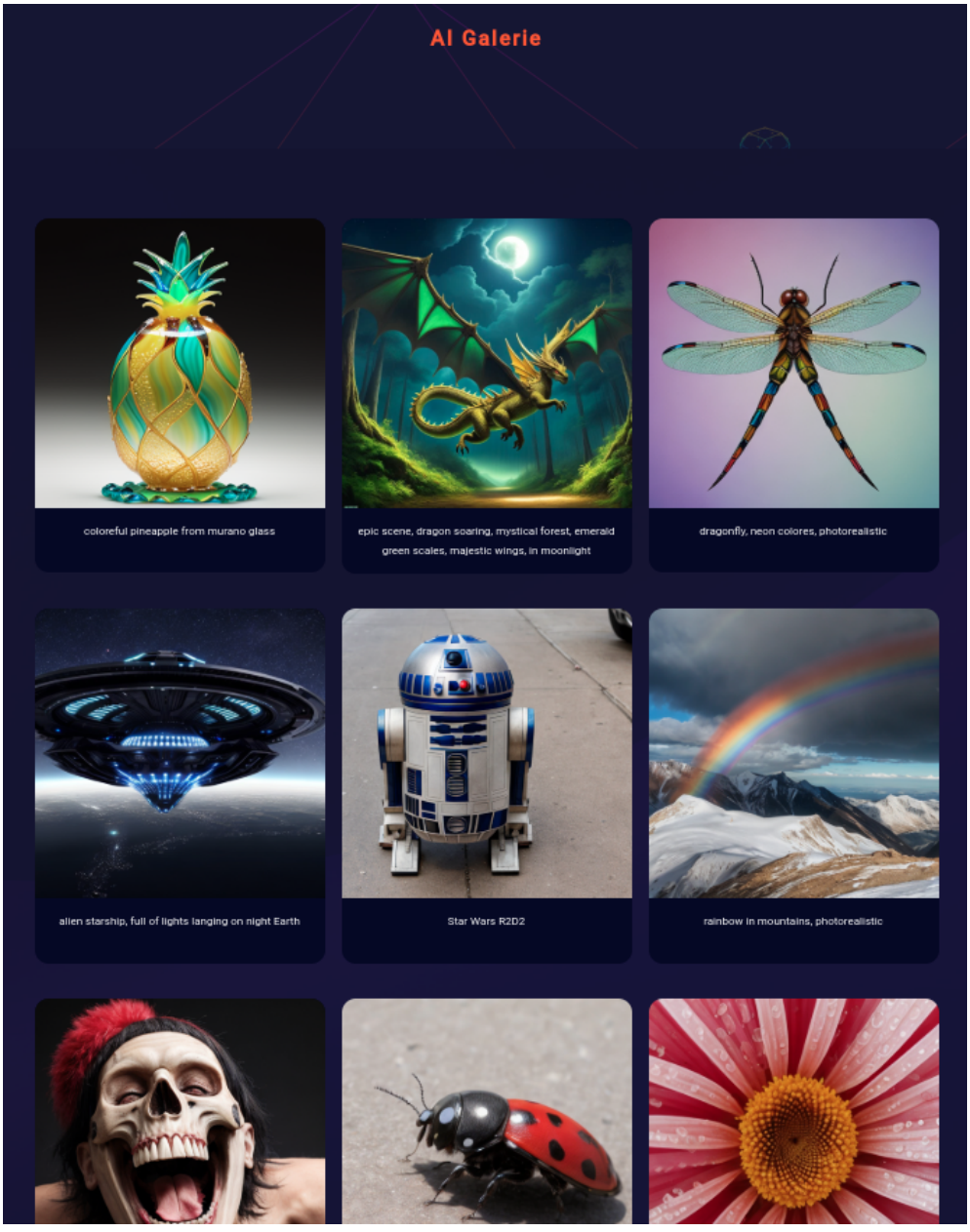
Obrázek 27: Webová nabídka videoprůvodců na webu



Obrázek 28: Stránka s videoprůvodcem instalací

4.2.3 Galerie praktických ukázek

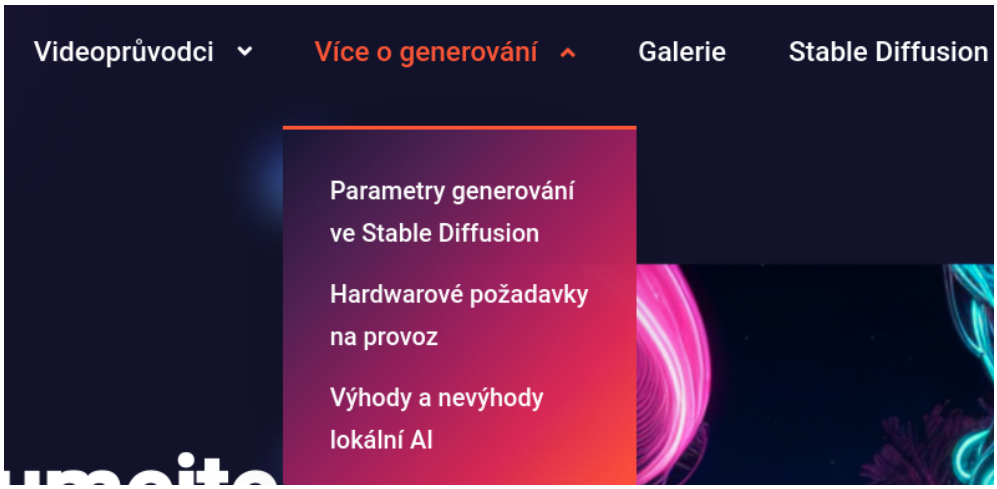
Pro lepší pochopení plného potenciálu Stable Diffusion jsem na webu vytvořila pestrou kolekci odlišných vygenerovaných obrázků.



Obrázek 29: Náhled na galerii webu

4.2.4 Více o problematice generování

Tato stať podrobněji přibližuji různé oblasti generování, přičemž je dáván důraz na tři klíčové oblasti. První část se věnuje parametrům generování, kde se detailně popisuje, jak jednotlivá nastavení, jako je rozlišení, počet kroků nebo třeba jak modely ovlivňují kvalitu a detaily výsledného obrazu. Další část se zaměřuje na vliv hardwaru, konkrétně jak různé konfigurace zařízení, jako jsou grafické karty nebo procesory, mohou ovlivnit efektivitu a rychlost generování. Cílem je pochopit, jaký hardware použít a jak jej optimálně nakonfigurovat pro dosažení co nejlepších výsledků. V poslední části se analyzují výhody modelu Stable Diffusion, který je jedním z nejpopulárnějších generativních nástrojů, a ukazuje se, jak jeho flexibilita a možnosti přizpůsobení usnadňují tvorbu. Celkovým cílem těchto kapitol je poskytnout hlubší vhled do technologií a metod, které stojí za generováním obrázků, a pomoci uživatelům lépe porozumět všem faktorům, které ovlivňují požadované výsledky.



Obrázek 30: Webová nabídka s kapitolami o generování

4.3 Tvorba webu

Pro vytvoření interaktivního průvodce generováním obrázků pomocí modelu EasyDiffusion jsem navrhla webovou prezentaci.

Účel úvodní strany je navigace při prvním kontaktu. Seznamuje návštěvníky s tím, co mohou na webu nalézt. Poskytuje rychlý přístup k hlavním sekcím obsahu, jako jsou slovník, galerie příkladových vygenerovaných obrázků, videoprůvodci pro instalaci a použití Stable Diffusion, a další stránky s podrobnějšími popisy k vlastnímu generování obrázků za pomoci umělé inteligence.

Slovník poskytuje přehled základních termínů, které se vztahují ke generování obrázků pomocí AI. Návštěvník webu zde nalezne definice, které mu pomohou lépe porozumět technologiím, procesům a metodám, jež stojí za tvorbou výstupů prostřednictvím generativních modelů jako je Stable Diffusion.

Sekce s videoprůvodci obsahuje návody, které provádějí uživatele instalací a používáním Stable Diffusion, včetně specifikací pro různé operační systémy a tipů pro efektivní generování obrázků pomocí specifikovaných promptů. Videonávody jsou doplněny o doprovodné texty, které detailně vysvětlují postupy krok za krokem.

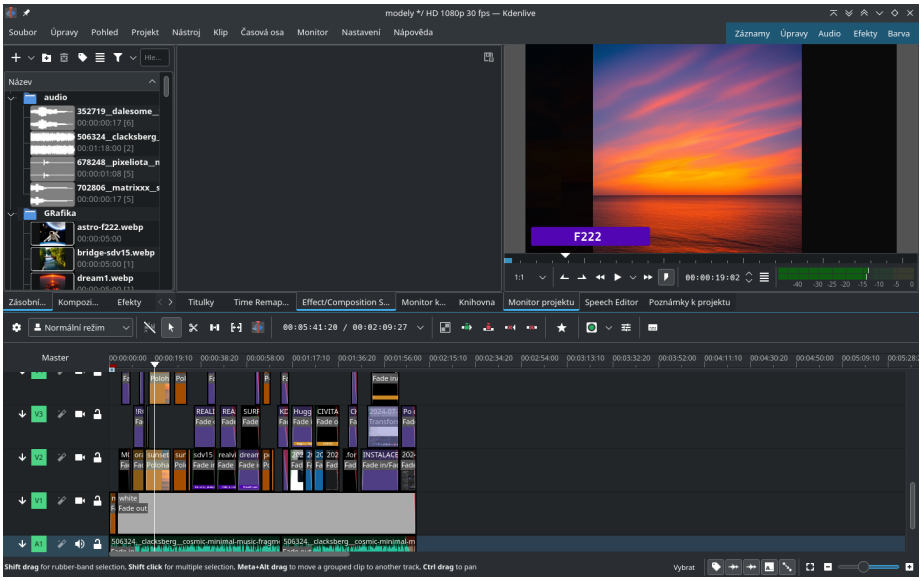
Stránky s podrobnějšími popisy obsahují informace o parametrech, vlivu hardwaru a výhodách Stable Diffusion, ovlivňujících generování obrázků pomocí AI. Návštěvník získá přehled o dalších aspektech, které formují výsledky generativních modelů.

Pro lepší pochopení plného potenciálu Stable Diffusion je součástí webu galerie obsahující širokou škálu odlišných a pestrých obrázků, které byly vygenerovány právě prostřednictvím této aplikace.

Pro vytvoření webu jsem použila kombinaci technologií HTML, CSS, frameworků Bootstrap5 a jQuery a několika dalších javascriptových komponent, jako je např. Owl Carousel či Swiper pro vytváření animovaných karuselových prezentací. Pomocí HTML tagů jsem definovala strukturu jednotlivých prvků, CSS zajišťuje webu vzhled, Bootstrap mu poskytuje předpřipravené komponenty a jQuery umožňuje jednoduché ovládaní HTML prvků přes HTML DOM.

4.4 Tvorba videoprůvodců

Abych vytvořila srozumitelné návody, musela jsem pečlivě připravit veškeré potřebné materiály. Záznamy obrazovky, které jsem pořizovala v programu OBS Studio, statické screenshoty a vlastní grafiku. Zvukové stopy byly nahrávány odděleně a následně synchronizovány se záznamem obrazovky. Komentář jsem přednášela s důrazem na jasné a srozumitelné předání informací, aby i méně zkušení uživatelé mohli snadno sledovat jednotlivé kroky. Všechny materiály byly následně sestaveny a upraveny do finální podoby videa v programu Kdenlive. Cílem bylo vytvořit sérii videonávodů, které by komplexně pokryly celý proces generování obrázků pomocí generativního modelu Stable Diffusion a poskytly uživatelům dostatek inspirace pro vlastní tvorbu. Součástí každého videoprůvodce je také doprovodný text umístěný na webu.



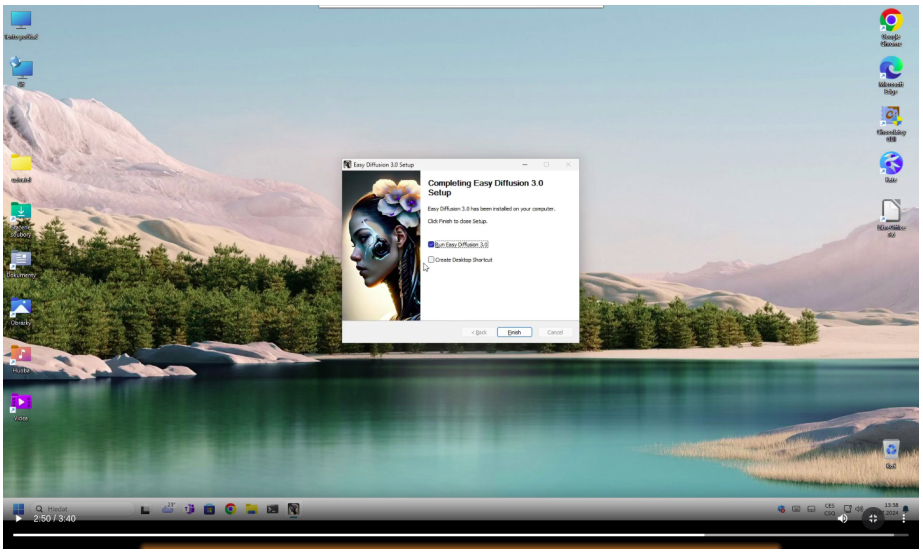
Obrázek 31: Střih videonávodu na téma modely v Kdenlive

4.4.1 Jak nainstalovat Stable Diffusion pomocí EasyDiffusion

V tomto videu uživatele navádím krok za krokem, jak nainstalovat Stable Diffusion s využitím instalátoru EasyDiffusion. Návod zahrnuje stažení potřebného softwaru, instalaci a jeho základní konfiguraci. Postup je uveden jak pro Linux a macOS, tak pro Windows.

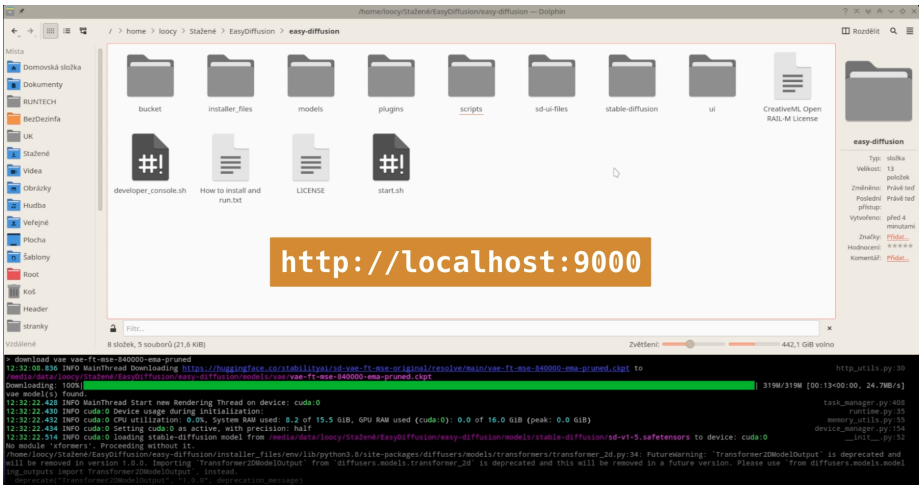
Postup instalace shrnutý ve videu

- Stažení instalátoru pro Linux / macOS / Windows na URL adrese https://easydiffusion.github.io/docs/installation
- Spuštění instalátoru
  - Linux a macOS: Spustit skript ./start.sh v terminálu
  - Windows: Spustit stažený .exe soubor



Obrázek 32: Snímek z videonávodu instalací Windows

- Po instalaci se AI rovnou spustí jako webová služba na portu 9000 (možno později nastavit na libovolný volný), proto stačí ve webovém prohlížeči přejít na adresu http://localhost:9000 (pokud spouštíte prohlížeč na stejném počítači, na kterém máte spuštěnu AI) nebo http://NAZEV_PC_NEBO_IP_VE_VASI_SITI:9000 (pokud přistupujete z jiného počítače ve vaší síti).



Obrázek 33: Snímek z videoprůvodce instalací Linux

- V praxi je tedy příhodné spustit prostředí na počítači s dostačujícím hardware a následně je možné přejít na webovou adresu např. na mobilu či tabletu. Využíváte tedy výkon svého počítače a parametry generování při tom zadáváte na mobilním zařízení.

- Pro opětovné spuštění ve Windows stačí spustit EasyDiffusion přes ikonu zástupce, v Linuxu / macOS opět přes spuštění skriptu start.sh v terminálu, nebo je možné si vytvořit spouštěcí ikonu jako je ve Windows po instalaci.

4.4.2 Princip generování obrázků ve Stable Diffusion

Video znázorňuje princip generování obrázků pomocí generativního modelu Stable Diffusion, především odstraňováním obrazového šumu. Ukáže uživateli jak model vytváří obrázky z textových popisů a jak funguje celý proces v pozadí. Doprovodný text následně problematiku popisuje více do hloubky.

V rámci tohoto tématu se uživatel dozví následující informace...

Generování obrázků s využitím AI je komplexní a vícefázový proces, který integruje různé technologie strojového a hlubokého učení. Základní princip spočívá v tom, že AI model, který byl trénován na rozsáhlém množství dat, dokáže na základě zadaného promptu vytvořit nový obrázek.

Mezi nejčastěji používané modely pro generování obrázků patří GAN (Generative Adversarial Networks) a difúzní modely. Pro účely tohoto popisu se zaměříme na princip difuzního modelu konkrétně na model latentní difúze, vzhledem k tomu, že práce se věnuje generátoru obrázků Stable Diffusion, který je populární implementací tohoto přístupu.

Latentní difúze je také důvod, proč Stable Diffusion dokáže generovat obrázky s vyšším rozlišením tak rychle – proces difúze není spouštěn na samotných pixelových obrázcích, ale na komprimovaném obraze v latentním prostoru.

Generování obrázků lze rozdělit do tří kroků.

Zadání promptu a ostatních parametrů

Prvním krokem v procesu generování obrázku je vyplnění parametrů pro generování včetně promptu, tedy jak již bylo zmíněno, textového popisu, který specifikuje, jak by měl výsledný obrázek vypadat. Tento krok je zásadní, protože poskytuje modelu informace o tom, co má být na obrázku zobrazeno.

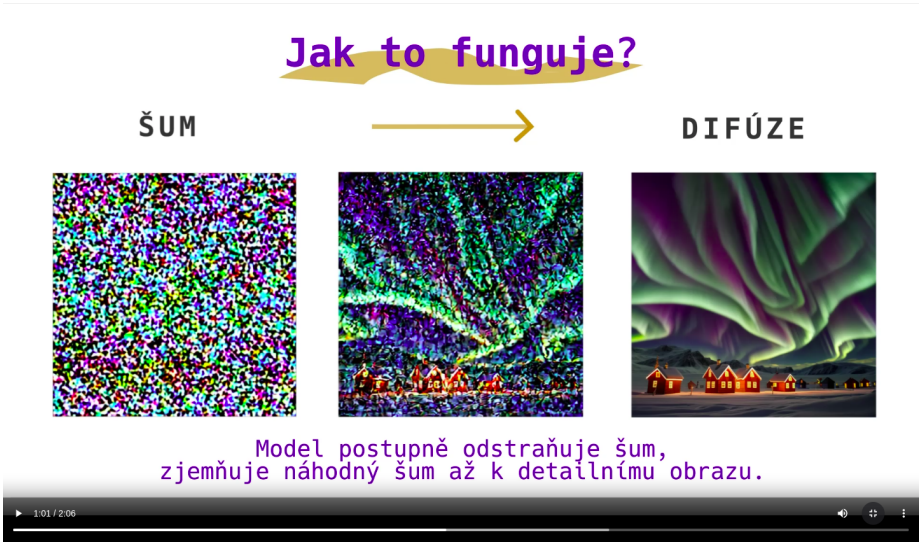
Předzpracování promptu

Algoritmus nejprve analyzuje zadaný prompt.

- Tokenizace
  - Výzva je rozdělena na jednotlivá slova, fráze nebo dokonce části slov. Například prompt fastfood shop by byl rozdělen na 3 tokeny: fast, food a shop. Tokeny jsou číselné reprezentace „slov“, které model může zpracovat.
  - Maximální množství tokenů na prompt je 77, respektive 75, jelikož ke každému promptu je přidán startovací a ukončovací token.
- Embeding (vkládání)
  - Každý token se převede na jeho vektorovou reprezentaci. Vektor zachycuje význam slova v kontextu promptu.
- Cross-attention (křížová pozornost)
  - Jde o klíčovou metodu Stable Diffusion, která umožňuje efektivní propojení textových promptů s generovanými obrazy.
  - Funguje tak, že model udržuje mapu vztahů mezi slovy a částmi obrazu, což umožňuje efektivní a přesné generování obrázků na základě textových popisů. Tento mechanismus je opakovaně využíván v celém procesu generování, čímž se zvyšuje přesnost a kvalita výsledného obrázku.
  - Výstup je využíván šumovým prediktorem v celém U-Netu (což je je typ konvoluční neuronové sítě). U-Net "čte" a "reaguje" na textový prompt pro vytváření kontextově relevantních obrázků.

Krokování samotného procesu generování

- Model generuje náhodný tenzor[2] v latentním prostoru, který slouží jako výchozí bod pro generování obrazu. Tento proces je řízen seedem generátoru náhodných čísel, který zajišťuje, že při stejném seedu bude generován stejný náhodný tenzor.
- Variační Automatický Kodér (VAE) převede vstupní obraz (tenzor) do latentního prostoru.
- Prediktor šumu odhaduje šum a následně ho odstraňuje z latentního tenzoru. Tento proces se opakuje v několika iteracích, kdy každá iterace zmenšuje množství šumu.
- Po odstranění šumu z latentního tenzoru dekodér VAE převádí latentní reprezentaci zpět do původního obrazového prostoru.



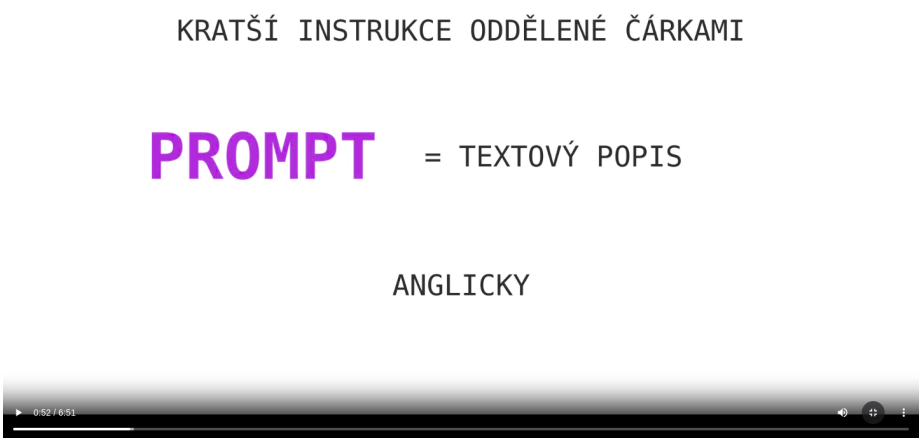
Obrázek 34: Snímek z videoprůvodce principu generování

2 Tenzor je více-dimenzionální pole čísel, které reprezentuje data. Když Stable Diffusion generuje náhodný tenzor v latentním prostoru, znamená to, že vytvoří náhodnou sadu čísel, která slouží jako výchozí bod pro generování obrazu.

4.4.3 Jak na prompty pro Stable Diffusion

Toto video je zaměřené na syntaxi promptů, vysvětluji v něm mimo jiné i funkci variant a vážení promptu. Ukazuji jak různé prompty ovlivňují výsledné obrázky.

Prompt, neboli instrukce v textové podobě, se pro Stable Diffusion zadávají v angličtině. Správně strukturované prompty umožňují modelu lépe porozumět požadavkům uživatele a vytvořit odpovídající výstup.

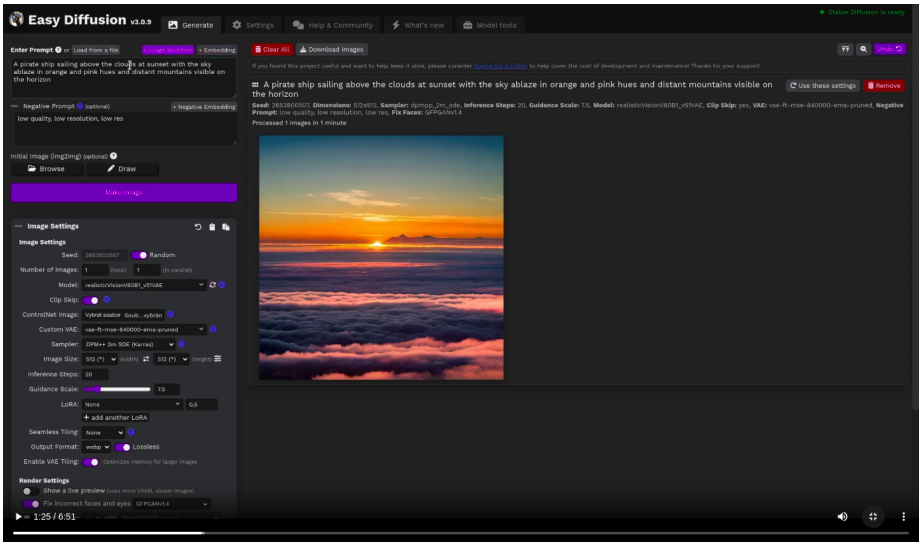


Obrázek 35: Snímek z videoprůvodce tvorby promptů

V průvodci jsem kladla důraz na to, že než psaní celých vět jsou efektivnější kratší instrukce oddělené čárkami. Model tak lépe chápe požadavky a interpretace je většinou přesnější.

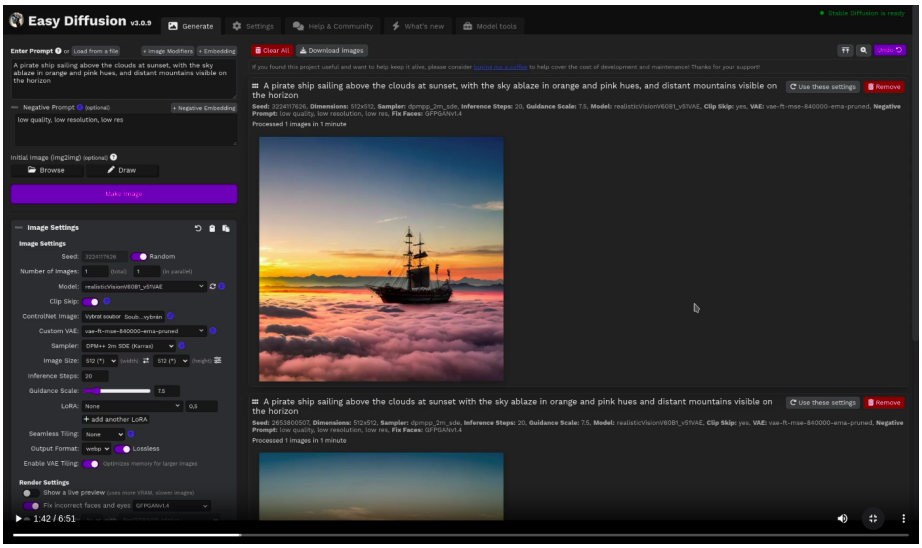
Ve videu to ukazuji na promptu, který je při prvním generování tvořen větou a při dalším generování je rozdělen na kratší části oddělené čárkami. První obrázek vůbec neobsahuje loď, která je hlavním prvkem zadání a díky tomu ho nesplňuje. Pouhým rozdělení promptu čárkami jsem docílila druhého obrázku, který naprosto odpovídá textovému popisu.

Nevhodná stavba promptu: a pirate ship sailing above the clouds at sunset with the sky ablaze in orange and pink hues and distant mountains visible on the horizont.



Obrázek 36: Snímek průběhu generování - nevyhovující syntaxe promptu

Vhodná stavba promptu: a pirate ship sailing above the clouds at sunset, with the sky ablaze in orange and pink hues, and distant mountains visible on the horizont.



Obrázek 37: Snímek průběhu generování - vyhovující syntaxe promptu

Varianty promptu

Dále uživatele seznamuji s možností různých variant promptu. V jednom promptu totiž můžeme testovat různé možnosti a variace. To je užitečné pro experimentování s různými styly nebo prvky. Jednotlivé varianty se uvádí do složených závorek.

Prompt photography of {young, old} woman with red hair, high quality ve skutečnosti zařadí do fronty generování 2 výstupů. A to pro prompt photography of young woman with red hair, high quality a photography of old woman with red hair, high quality.

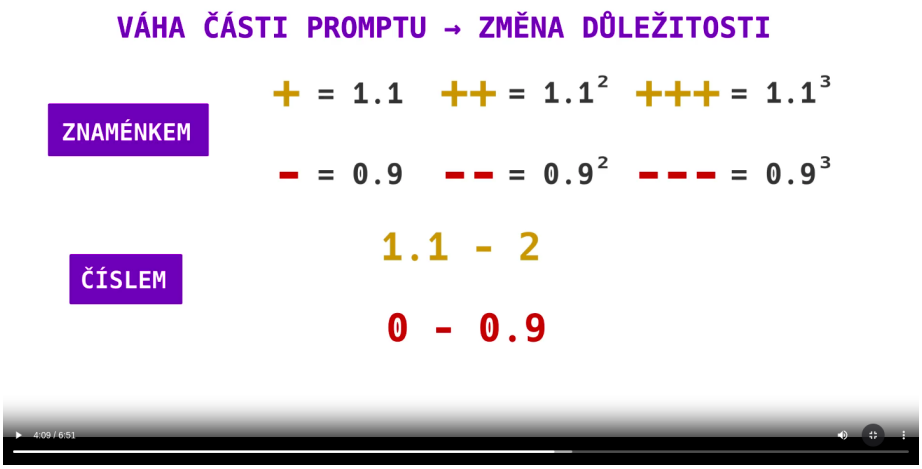
Váha promptu

Ve videu jsem představila také možnost nastavit váhu jednotlivých částí promptu, což generativní model Stable Diffusion umožňuje. Váhy zvýrazní určité aspekty a zadáváme je vždy za konkrétní část promptu. Ve videu jsem shrnula následující.

Výchozí váha jednotlivých frází je hodnota 1.

Pokud chceme zvýšit důležitost, použijeme znaménko + nebo číslo od 1.1 do 2. Platí, že + je ekvivalentem pro 1.1, ++ je totéž co 1.1^2 a tak dále.

Potřebujeme li snížit důležitost, využijeme znaménko – nebo čísla od 0 do 0.9. Zde obdobně platí, že – je rovno 0.9, -- má stejný účinek jako 0.9^2 a stejný princip je uplatňován i pro následující hodnoty.



Obrázek 38: Snímek z videoprůvodce jak na prompty - váhy

Pokud potřebujeme dodat váhu pouze jednomu aspektu promptu, provedeme to následovně:

- Znaménkem: street artist creating a graffiti+++, vibrant colors, urban setting
- Číselnou hodnotou: street artist creating (a graffiti)1.3, vibrant colors, urban setting

Při vážení více frází popisu je nutné všechny tyto části uzavřít do závorek:

- Se znaménky: (street)- artist creating (a graffiti)+++, vibrant colors
- S číselnými hodnotami: (street)0.9 artist creating (a graffiti)1.3, vibrant colors

Váhy je možné do sebe nořit. Váha (grafiti+)+ odpovídá grafiti++.

Hory hor, tak by se dala pojmenovat tato ukázka vážení promptu. Základní prompt zní beautiful landscape, mountains, rivers, trees, sunset.



Obrázek 39: Vygenerovaný obrázek po zadání promptu "beautiful landscape, mountains, rivers, trees, sunset"

Váha vyšší i nižší byla přidávána na část „mountains“. Prompty tedy zněly následovně:

- beautiful landscape, mountains+++, rivers, trees, sunset
- beautiful landscape, mountains---, rivers, trees, sunset

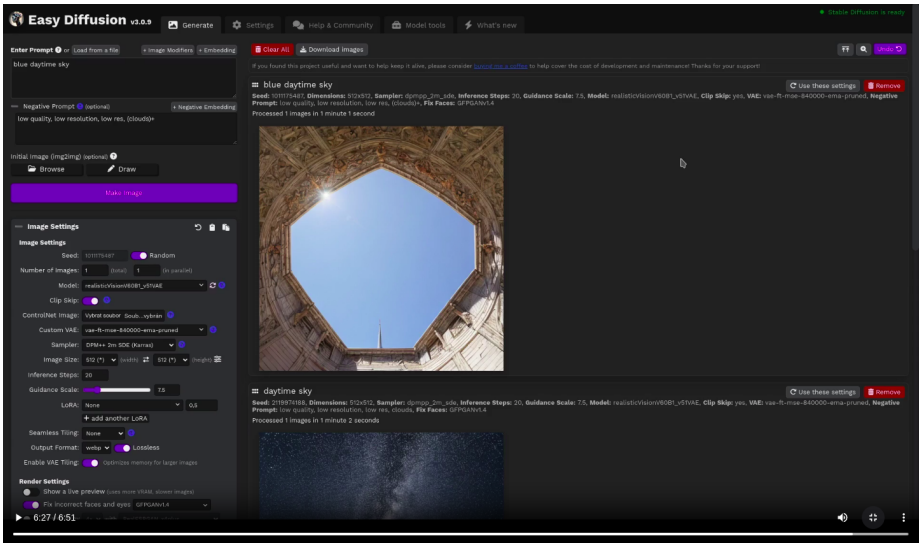


Obrázek 40: Obrázek vygenerovaný po zadání promptu "beautiful landscape, (mountains)+++, rivers, trees, sunset"



Obrázek 41: Obrázek vygenerovaný po zadání promptu "beautiful landscape, (mountains)---, rivers, trees, sunset"

Co na obrázku chceme, sdělíme v promptu. Co když naopak na obrázku něco nechceme? K tomu slouží negativní prompty a i o těch je toto video. Jako ukázku jsem nechala vygenerovat denní oblohu bez mraků. Jako negativní prompt, jsem tedy zadala cloud.



Obrázek 42: Snímek z videoprůvodce jak na prompty - negativní prompt

Mezi mnou často zadávané negativní prompty patří low resolution nebo low quality, hojně se také jako záporný příkaz používá slovo text.

4.4.4 Přehled a použití modelů v Stable Diffusion

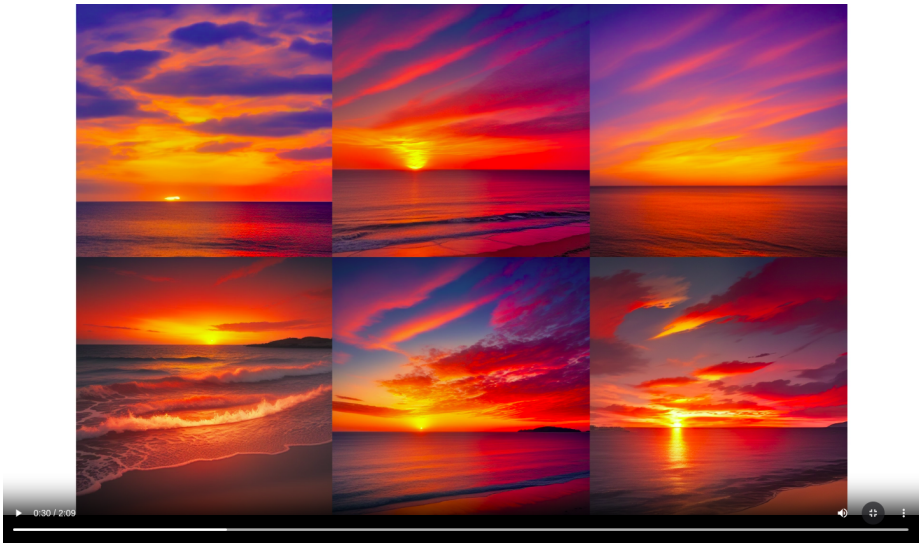
Video poskytuje přehled různých modelů dostupných v Stable Diffusion, vysvětluje jejich specifické vlastnosti a ukazuje, jak si vybrat správný model pro konkrétní úkol a jak vhodně využívat různé modely k dosažení požadovaných výstupů.

Výběr modelu je důležitý, jelikož udává styl výsledného obrázku. Nutno podotknout, že na styl jsou navázány ostatní parametry, jako je rozlišení. Při rozlišení 1024×1024 je už třeba použít XL verze modelů.

Ve videu blíže představuji následující modely:

- SD v1-5 – široká škála užití, realismus, vysoké detaily
- Realistic Vision – realistické zobrazení lidí, zvířat, krajin
- DreamShaper – abstraktní tvorba, fantasy stylizace

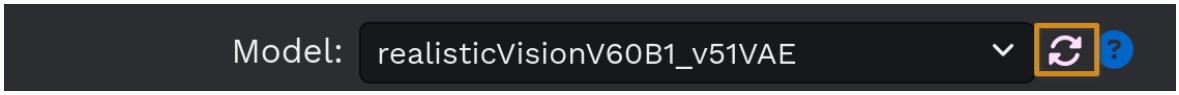
Pro čistě vizuální porovnání jsem přidala ještě modely Anything, Deliberate a F222.



Obrázek 43: Snímek z videoprůvodce modely

Stažení a instalace modelů pro Stable Diffusion

- Jednotlivé modely je možné vyhledat a stáhnout např. na webových stránkách CIVITAI či Hugging Face
- Modelový soubor se stáhne ve formátu .ckpt nebo .safetensors
- Stažený modelový soubor je třeba přesunout do složky models ve vašem adresáři s instalací EasyDiffusion
- V uživatelském rozhraní následně stačí obnovit seznam modelů



Obrázek 44: Ukázka obnovení modelů v EasyDiffusion

Závěr

Tato bakalářská práce se zaměřovala na vytvoření komplexního průvodce pro generování obrázků pomocí lokálně provozovaného modelu umělé inteligence Stable Diffusion, s využitím instalátoru EasyDiffusion. Hlavním cílem bylo poskytnout uživatelům, zejména studentům vysokých škol, efektivní nástroj pro práci s generativní umělou inteligencí, který zahrnuje jak teoretické znalosti, tak praktické dovednosti prostřednictvím videonávodů.

V teoretické části byly shrnuty základní principy a technologie umělé inteligence používané při generování obrázků. Bylo provedeno srovnání cloudových a lokálně provozovaných nástrojů, přičemž zvláštní pozornost byla věnována softwaru Stable Diffusion a jeho možnostem. Byly představeny klíčové pojmy jako difúzní model, latentní prostor, a byly vysvětleny důležité aspekty procesu generování obrázků.

Praktická část práce byla zaměřena na vytvoření webové stránky s videonávody, které uživatele provádí instalací a používáním Stable Diffusion a EasyDiffusion. Vytvořené videonávody pokrývají základní instalaci, principy generování obrázků, tvorbu promptů, přehled a použití modelů. Tyto videonávody byly navrženy tak, aby byly co nejvíce uživatelsky přívětivé a umožňovaly snadné osvojení technologie krok za krokem.

Výsledkem této práce je ucelený průvodce, který umožňuje uživatelům efektivně využívat technologii generování obrázků na lokálních zařízeních. Tento průvodce poskytuje nejen teoretické znalosti, ale i praktické dovednosti, čímž zajišťuje komplexní pochopení a využití umělé inteligence v oblasti tvorby grafiky. Práce přispívá k lepšímu porozumění a širšímu využívání generativní AI v lokálním prostředí, což může být přínosné jak pro akademické účely, tak pro osobní projekty uživatelů.

Seznam použitých informačních zdrojů

[1] *Co to vlastně je umělá inteligence. Online. AI DĚTEM, Z.S. AI dětem. Dostupné z: https://aidetem.cz/obecny-uvod-do-umele-inteligence/co-to-vlastne-je-ai. [cit. 2024-11-28].*

[2] *Slovníček. Online. AI DĚTEM, Z.S. AI dětem. Dostupné z: https://aidetem.cz/slovnicek-pojmu-umele-inteligence. [cit. 2024-07-08].*

[3] *What is a neural network? Online. IBM. Dostupné z: https://www.ibm.com/topics/neural-networks. [cit. 2024-07-08].*

[4] *What is machine learning? Online. SAP. Dostupné z: https://www.sap.com/products/artificial-intelligence/what-is-machine-learning.html. [cit. 2024-07-08].*

[5] LEARNOPENCV TEAM. *Generative and Discriminative Models*. Online. LearnOpenCV. 2021. Dostupné z: https://learnopencv.com/generative-and-discriminative-models. [cit. 2024-07-08].

[6] SINGH, Vaibhav a RATH, Sovit. *Diffusion Models for Image Generation – a Comprehensive Guide*. Online. LearnOpenCV. 2023. Dostupné z: https://learnopencv.com/image-generation-using-diffusion-models. [cit. 2024-07-08].

[7] Stable Diffusion. Online. © 2024. Dostupné z: https://stablediffusionweb.com. [cit. 2024-11-28].

[8] EasyDiffusion. Online. © 2024. Dostupné z: https://easydiffusion.github.io. [cit. 2024-11-28].

[9] CMDRGRUNT. *Writing prompts*. Online. GITHUB, INC. GitHub. 2024. Dostupné z: https://github.com/easydiffusion/easydiffusion/wiki/Writing-prompts. [cit. 2024-07-08].

[10] CMDRGRUNT. *Image Modifiers*. Online. GITHUB, INC. GitHub. 2024. Dostupné z: https://github.com/easydiffusion/easydiffusion/wiki/Image-Modifiers [cit. 2024-07-08].

[11] CMDR2. *ControlNet*. Online. GITHUB, INC. GitHub. 2023. Dostupné z: https://github.com/easydiffusion/easydiffusion/wiki/ControlNet. [cit. 2024-07-08].

[12] JELUF. *VAE Variational Auto Encoder*. Online. GITHUB, INC. GitHub. 2022. Dostupné z: https://github.com/easydiffusion/easydiffusion/wiki/VAE-Variational-Auto-Encoder. [cit. 2024-07-08].

[13] CMDR2. *Samplers*. Online. GITHUB, INC. GitHub. 2023. Dostupné z: https://github.com/easydiffusion/easydiffusion/wiki/Samplers. [cit. 2024-07-08].

[14] CMDR2. *Seamless Tiling*. Online. GITHUB, INC. GitHub. 2023. Dostupné z: https://github.com/easydiffusion/easydiffusion/wiki/Seamless-Tiling. [cit. 2024-07-08].

Seznam obrázků

Obrázek 1: Vygenerovaný obrázkem promptem "closeup photo of brown cat in the mug which is in the kitchen lit by daylight".....17

Obrázek 2: Vygenerovaný obrázek po zadání promptu "brown cat in the mug, closeup photo, kitchen background, dayligth".....18

Obrázek 3: Ukázka modifikátorů obrázku v EasyDiffusion.....19

Obrázek 4: Ukázka parametru Seed v EasyDiffusion.....19

Obrázek 5: Ukázka parametru Number of Images v EasyDiffusion.....20

Obrázek 6: Ukázka generování 2 obrázku najednou v EasyDiffusion.....20

Obrázek 7: Vygenerovaný obrázek po zadání promptu "realistic illustration of heart organ, detailed sketch, human anatomy".....21

Obrázek 8: Vygenerovaný obrázek po zadání promptu "photography of underwater scenery+ +, glowing corals, fishs, Colorful".....22

Obrázek 9: Vygenerovaný obrázek po zadání promptu "high resolution photography interior design, a cozy scandinavian living room, wooden floor, many details, studio light".....23

Obrázek 10: Vygenerovaný obrázek po zadání promptu "a woman with dark hair, unforgettable, high quality".....24

Obrázek 11: Vygenerovaný obrázek po zadání promptu "a small balcone with the Eiffle Tower view, morning sun, amazing breakfast, high qualit".....25

Obrázek 12: Vygenerovaný obrázek po zadání promptu "photography of big (crocodile)+ in water, high quality".....26

Obrázek 13: Vygenerovaný obrázek po zadání promptu "oil painting of a future city of (flying islands)+, fantasy sci - fi city, fantasy style, highly detailed, magnificent".....27

Obrázek 14: Vygenerovaný obrázek po zadání promptu "a 3D model of one cute raspberry fruit, big smile, sharp teath, clean background".....28

Obrázek 15: Ukázka nastavení funkce ControlNet Image v EasyDiffusion.....29

Obrázek 16: Vygenerovaný obrázek prostřednictvím funkce ControlNet - JohnnyDepp…30
Obrázek 17: Ukázka vyplnění parametru VAE v EasyDiffusion.....30
Obrázek 18: Ukázka vyplnění parametru Sampler v EasyDiffusion.....31
Obrázek 19: Ukázka vyplnění parametru Image Size v EasyDiffusion.....31
Obrázek 20: Ukázka vyplnění parametru Inference Steps v EasyDiffusion.....31
Obrázek 21: Ukázka vyplnění parametru Guidance Scale v EasyDiffusion.....32
Obrázek 22: Navigace na možnost Settings v EasyDiffusion.....33
Obrázek 23: Ukázka možnosti přepnutí akcelerovaného běhu na běh na procesoru.....33
Obrázek 24: Úvodní stránka webových stránek.....42
Obrázek 25: Rozcestník webových stránek.....44
Obrázek 26: Stránka webu se slovníkem.....44
Obrázek 27: Webová nabídka videoprůvodců na webu.....45
Obrázek 28: Stránka s videoprůvodcem instalací.....45
Obrázek 29: Náhled na galerii webu.....46
Obrázek 30: Webová nabídka s kapitolami o generování.....47
Obrázek 31: Střih videonávodu na téma modely v Kdenlive.....49
Obrázek 32: Snímek z videonávodu instalací Windows.....50
Obrázek 33: Snímek z videoprůvodce instalací Linux.....51
Obrázek 34: Snímek z videoprůvodce principu generování.....54
Obrázek 35: Snímek z videoprůvodce tvorby promptů.....55
Obrázek 36: Snímek průběhu generování - nevyhovující syntaxe promptu.....56
Obrázek 37: Snímek průběhu generování - vyhovující syntaxe promptu.....56
Obrázek 38: Snímek z videoprůvodce jak na prompty - váhy.....57

Obrázek 39: Vygenerovaný obrázek po zadání promptu "beautiful landscape, mountains, rivers, trees, sunset"....58

Obrázek 40: Obrázek vygenerovaný po zadání promptu "beautiful landscape, (mountains)++ +, rivers, trees, sunset"....59

Obrázek 41: Obrázek vygenerovaný po zadání promptu "beautiful landscape, (mountains)---, rivers, trees, sunset"....59

Obrázek 42: Snímek z videoprůvodce jak na prompty - negativní prompt....60

Obrázek 43: Snímek z videoprůvodce modely....61

Obrázek 44: Ukázka obnovení modelů v EasyDiffusion....62