

---

## Efektivita difúzných modelov v multimedialnej tvorbe

### The effectiveness of diffusion models in multimedia creation

Peter Procházka<sup>1</sup>

#### Abstrakt

Tento výskum sa zameriava na analýzu efektivity generatívnych difúzných modelov v kontexte multimedialnej tvorby. Skúmame vzťah medzi nastavením parametrov týchto modelov, kvalitou generovaného obsahu a výpočtovou náročnosťou. Prostredníctvom série experimentov hodnotíme vplyv rôznych hyperparametrov na kvalitu výstupu a analyzujeme kompromis medzi kvalitou a výpočtovou efektivitou. Naše zistenia poskytujú cenné poznatky pre optimalizáciu využitia difúzných modelov v reálnych aplikáciách a naznačujú smer pre budúci výskum v oblasti generatívnych modelov pre multimedialnu tvorbu.

#### Kľúčové slová

difúzne modely, generatívne modelovanie, multimedialna tvorba, výpočtová efektivita, kvalita obrazu, umelá inteligencia

#### Abstract

This research focuses on analyzing the effectiveness of generative diffusion models in the context of multimedia content creation. We examine the relationship between the parameter settings of these models, the quality of generated content, and computational complexity. Through a series of experiments, we evaluate the impact of various hyperparameters on output quality and analyze the trade-off between quality and computational efficiency. Our findings provide valuable insights for optimizing the use of diffusion models in real-world applications and suggest directions for future research in the field of generative models for multimedia content creation.

#### Key words

diffusion models, generative modeling, multimedia creation, computational efficiency, image quality, artificial intelligence

#### JEL classification

C45, C61, C63

## 1 Úvod

V posledných rokoch sme svedkami revolúcie v oblasti umelej inteligencie, pričom generatívne modely založené na difúzných procesoch zaujali popredné miesto v generovaní multimedialného obsahu. Tieto modely prinášajú nové možnosti v oblasti počítačového videnia, spracovania prirodzeného jazyka a tvorby multimedialného obsahu (Dhariwal a Nichol, 2021). Ich schopnosť vytvárať realistické a vysoko kvalitné výstupy otvorila nové horizonty v rôznych odvetviach, od umenia a dizajnu až po vedecký výskum a vývoj produktov.

Difúzne modely, ktoré boli pôvodne predstavené Sohlem-Dicksteinom et al. (2015), fungujú na princípe postupného pridávania šumu do dát a následného učenia sa tento proces

---

<sup>1</sup> Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky, Katedra aplikovanej informatiky, Dolnozemska cesta 1, 852 35 Bratislava, peter.prochazka@euba.sk.

zvrátiť. Tento prístup sa ukázal byť mimoriadne efektívny pri generovaní rôznych typov multimedialného obsahu, vrátane obrazov, zvukov a videí (Ho et al., 2020). Napriek ich nepopierateľnému potenciálu však zostáva mnoho otázok ohľadom optimálneho nastavenia parametrov týchto modelov, ich výpočtovej efektivity a kvality generovaného obsahu.

V kontexte multimedialných aplikácií sú generatívne modely založené na difúzných procesoch obzvlášť zaujímavé vďaka ich schopnosti zachytiť komplexné štatistické vzory v dátach. Ramesh et al. (2022) demonštrovali ich potenciál v oblasti generovania obrazov na základe textových popisov, čo otvorilo nové možnosti pre interaktívnu tvorbu obsahu. Zároveň však Song et al. (2021) poukázali na výzvy spojené s kontrolou a interpretáciou výstupov týchto modelov, čo zdôrazňuje potrebu ďalšieho výskumu v tejto oblasti.

V tomto článku sa zameriavame na analýzu použitia generatívnych modelov založených na difúzných procesoch v multimedialných aplikáciách. Naším cieľom je preskúmať, ako rôzne parametre týchto modelov ovplyvňujú kvalitu generovaného obsahu a aký je vzťah medzi výpočtovou náročnosťou a kvalitou výstupu. Taktiež sa zameriame na porovnanie týchto modelov s existujúcimi prístupmi v oblasti generovania multimedialného obsahu.

## 2 Ciele výskumu

Hlavným cieľom našej štúdie je vyhodnotiť efektivitu generatívnych difúzných modelov v multimedialnej tvorbe so zameraním na kvalitu výstupu a výpočtovú náročnosť. Tento cieľ sledujeme v kontexte rastúceho významu automatizovanej tvorby multimedialného obsahu a potreby efektívnych nástrojov v tejto oblasti (Rombach et al., 2022).

Na dosiahnutie nášho hlavného cieľa sme formulovali nasledujúce špecifické výskumné otázky:

1. Ako ovplyvňujú rôzne parametre difúzných modelov kvalitu generovaného multimedialného obsahu?
2. Aký je vzťah medzi výpočtovou náročnosťou a kvalitou výstupu pri použití týchto modelov?
3. Aké sú hlavné rozdiely medzi navrhovanou metódou a súčasnými postupmi pri tvorbe multimedialného obsahu?

Prostredníctvom týchto výskumných otázok sa snažíme nielen prispieť k teoretickému pochopeniu difúzných modelov, ale aj poskytnúť praktické poznatky pre ich efektívne využitie v multimedialných aplikáciách.

## 3 Teoretické základy a súvisiace práce

Generatívne modely predstavujú kľúčovú oblasť strojového učenia, ktorá sa zameriava na vytváranie nových dát na základe naučených vzorov. V rámci tohto širšieho kontextu, difúzne modely ponúkajú inovatívny prístup založený na princípoch postupnej transformácie dát. Tieto modely využívajú koncepty z teórie Markovových reťazcov a stochastických diferenciálnych rovníc na modelovanie komplexných distribúcií dát. Kľúčovým aspektom je tzv. score-based generovanie, ktoré umožňuje efektívne vzorkovanie z týchto distribúcií.

### 3.1 Generatívne modely

Oblasť generatívnych modelov v strojovom učení zaznamenala v posledných rokoch výrazný pokrok. Tieto modely sa snažia zachytiť podstatné charakteristiky tréningových dát a využiť ich na tvorbu nového, syntetického obsahu. Na rozdiel od diskriminačných modelov, ktoré sa zameriavajú na klasifikáciu alebo predikciu, generatívne modely sa učia reprezentovať samotnú distribúciu dát. Tieto modely sa v posledných rokoch stali predmetom intenzívneho výskumu vďaka ich schopnosti generovať realistický obsah v rôznych doménach, vrátane obrazov, textu a zvuku (Goodfellow et al., 2014).

Medzi populárne prístupy v oblasti generatívnych modelov patria variačné autoenkodéry (VAE), generatívne adversariálne siete (GAN) a v poslednej dobe aj autoregresívne modely. Každý z týchto prístupov má svoje špecifické výhody a obmedzenia. Napríklad, GAN-y sú známe schopnosťou produkovať vysoko kvalitné výstupy, ale často trpia nestabilitou pri tréňovaní a módom kolapsom.

V tomto kontexte predstavujú difúzne modely nový a sľubný prístup, ktorý sa snaží prekonať niektoré z limitácií existujúcich metód. Ich unikátny prístup k modelovaniu distribúcie dát ponúka potenciál pre stabilnejšie tréňovanie a lepšiu kontrolu nad procesom generovania.

### 3.2 Difúzne procesy v generatívnych modeloch

Difúzne modely predstavujú inovatívny prístup v oblasti generatívnych modelov. Ich základný princíp spočíva v postupnom pridávaní šumu do dát a následnom učení sa tento proces zvrátiť. Sohl-Dickstein et al. (2015) predstavili tento koncept ako alternatívu k tradičným generatívnym modelom, pričom poukázali na jeho potenciál pre vytváranie vysoko kvalitných výstupov.

Kľúčovým aspektom difúzných modelov je ich schopnosť modelovať komplexné distribúcie dát prostredníctvom série jednoduchších transformácií. Tento prístup umožňuje efektívne učenie a generovanie v priestoroch s vysokou dimenzionalitou, čo je obzvlášť výhodné pre multimedialne aplikácie (Ho et al., 2020).

### 3.3 Aplikácie v multimedialnej oblasti

Difúzne modely našli široké uplatnenie v multimedialnej oblasti. Ramesh et al. (2022) demonštrovali ich potenciál v generovaní obrazov na základe textových popisov, čím otvorili nové možnosti pre interaktívnu tvorbu vizuálneho obsahu. Okrem už spomenutého generovania obrazov na základe textových popisov, tieto modely preukázali schopnosť vykonávať komplexné úpravy existujúcich obrazov. Napríklad, Meng et al. (2021) predstavili metódu SDEdit, ktorá umožňuje sémantickú manipuláciu s obrazmi prostredníctvom difúzných modelov. Táto technika dovoľuje užívateľom vykonávať sofistikované úpravy, ako je zmena štýlu obrazu alebo pridávanie či odstraňovanie objektov, pričom zachováva vysokú úroveň realizmu.

V oblasti spracovania zvuku difúzne modely otvorili nové možnosti pre syntézu a konverziu reči. Kong et al. (2021) vo svojej práci o DiffWave ukázali, že tieto modely dokážu generovať vysoko kvalitné zvukové vlny s výnimočnou vernosťou. Tento prístup má potenciál revolucionizovať oblasti ako je syntéza reči pre asistenčné technológie alebo vytváranie realistických zvukových efektov pre filmový a herný priemysel.

Zaujímavou aplikáciou difúzných modelov je aj oblasť umeleckej tvorby. Briot et al. (2022) vo svojej štúdii skúmali využitie týchto modelov pre generovanie hudobných kompozícií. Zistili, že difúzne modely dokážu zachytiť komplexné štrukturálne vzťahy v hudbe, čo vedie k tvorbe koherentných a esteticky príjemných skladieb. Tento výskum otvára nové možnosti pre počítačom asistovanú hudobnú kompozíciu a interaktívne hudobné nástroje.

V oblasti video produkcie Ho et al. (2022) predstavili Video Diffusion Models (VDM), ktoré rozširujú koncept difúzných modelov na generovanie video sekvencií. Tento prístup umožňuje nielen vytváranie krátkych videí na základe textových popisov, ale aj interpoláciu medzi kľúčovými snímkami, čo má potenciálne aplikácie vo filmovej produkcii a vizuálnych efektoch.

Difúzne modely našli uplatnenie aj v oblasti rozšírenej a virtuálnej reality (AR/VR). Saito et al. (2023) vo svojom výskume využili tieto modely pre generovanie realistických 3D textúr a materiálov. Tento prístup významne zlepšuje vizuálnu kvalitu AR/VR prostredí a zároveň znižuje nároky na manuálnu prácu pri tvorbe 3D obsahov.

V neposlednom rade, difúzne modely prispievajú k pokroku v oblasti multimedialnej analýzy a porozumenia. Napríklad, Rombach et al. (2022) ukázali, že tieto modely môžu byť použité pre úlohy ako je sémantická segmentácia obrazov alebo detekcia objektov, čo rozširuje ich potenciálne využitie nad rámec čisto generatívnych úloh.

Tieto rozmanité aplikácie ilustrujú flexibilitu a potenciál difúzných modelov v multimedialnej oblasti. S pokračujúcim výskumom a vývojom očakávame, že sa spektrum aplikácií týchto modelov bude ďalej rozširovať, čo povedie k novým inovatívnym nástrojom a technikám v oblasti multimedialnej tvorby a spracovania.

### 3.4 Prehľad existujúcich prístupov a ich limitácií

Napriek značnému pokroku v oblasti difúzných modelov existujú stále výzvy, ktoré je potrebné riešiť. Song et al. (2021) poukázali na problémy spojené s kontrolou a interpretáciou výstupov týchto modelov, čo môže limitovať ich praktické využitie v niektorých aplikáciách. Nichol a Dhariwal (2021) zdôraznili dôležitosť správneho nastavenia hyperparametrov pre dosiahnutie optimálnych výsledkov, čo motivuje našu prvú výskumnú otázku.

## 4 Metodológia

Naša metodológia pre skúmanie efektivity difúzných modelov v multimedialnej tvorbe zahŕňa komplexný prístup k návrhu experimentov, výberu datasetov, implementácii modelov a hodnoteniu výsledkov. Zameriavame sa na analýzu vplyvu rôznych parametrov na kvalitu generovaného obsahu a výpočtovú náročnosť. Tento prístup nám umožňuje systematicky skúmať výkonnosť difúzných modelov v rôznych podmienkach a porovnávať ich s existujúcimi metódami.

### 4.1 Návrh experimentov

Na zodpovedanie našich výskumných otázok sme navrhli sériu experimentov zameraných na analýzu vplyvu rôznych parametrov difúzných modelov na kvalitu generovaného obsahu a výpočtovú náročnosť. Naše experimenty zahŕňajú:

1. Analýzu vplyvu počtu krokov difúzie na kvalitu výstupu
2. Skúmanie vzťahu medzi veľkosťou modelu a kvalitou generovaného obsahu
3. Hodnotenie rôznych stratégií vzorkovania z hľadiska výpočtovej efektivity a kvality výstupu.

### 4.2 Výber a príprava datasetov

Pre naše experimenty sme využili dva široko používané datasety:

1. CIFAR-10 (Krizhevsky, 2009) pre generovanie obrazov s nízkou rozlišovacou schopnosťou
2. CelebA-HQ (Karras et al., 2018) pre generovanie vysoko kvalitných portrétov

Tieto datasety sme zvolili pre ich rozmanitosť a široké využitie v oblasti výskumu generatívnych modelov, čo umožňuje porovnanie našich výsledkov s existujúcimi prácami.

### 4.3 Implementácia modelov

Implementovali sme difúzny model založený na architektúre U-Net, ktorá sa ukázala ako efektívna v rôznych úlohách generovania obrazu (Ronneberger et al., 2015). Naš model sme trénovali s rôznymi konfiguráciami hyperparametrov, vrátane počtu krokov difúzie ( $T = 100, 500, 1000$ ) a veľkosti modelu (počet parametrov = 5M, 20M, 50M).

#### 4.4 Metriky hodnotenia kvality a efektivity

Pre hodnotenie kvality generovaného obsahu sme použili nasledujúce metriky:

1. Fréchetova inception distance (FID) (Heusel et al., 2017) pre meranie kvality a diverzity generovaných obrazov.
2. Inception Score (IS) (Salimans et al., 2016) pre hodnotenie realistikosti a variability generovaných obrazov.

Pre hodnotenie výpočtovej efektivity sme merali:

1. Čas potrebný na generovanie jedného obrazu.
2. Využitie pamäte počas generovania.
3. Počet FLOPS (floating point operations per second) potrebných na generovanie obrazu.

### 5 Experimenty a výsledky

Naše experimenty sú navrhnuté tak, aby poskytli komplexný pohľad na výkonnosť difúzných modelov v kontexte multimedialnej tvorby. Zameriavame sa na tri kľúčové aspekty: vplyv parametrov modelu na kvalitu výstupu, analýzu výpočtovej náročnosti a porovnanie s existujúcimi prístupmi. Na hodnotenie kvality generovaného obsahu využívame kombináciu štandardných metrik a nových prístupov, ktoré zohľadňujú aj aspekty ľudského vnímania.

#### 5.1 Experiment 1: Vplyv parametrov na kvalitu výstupu

V prvom experimente sme skúmali vplyv počtu krokov difúzie ( $T$ ) na kvalitu generovaného obsahu. Zistili sme, že zvyšovanie  $T$  viedlo k zlepšeniu kvality výstupu, ale s klesajúcim prínosom pri vyšších hodnotách. Konkrétne (model s 50M parametrami):

- $T = 100$ : FID = 28.5, IS = 6.8
- $T = 500$ : FID = 24.7, IS = 7.3
- $T = 1000$ : FID = 22.9, IS = 7.6

Tieto výsledky naznačujú, že existuje bod, za ktorým ďalšie zvyšovanie  $T$  prináša len marginálne zlepšenie kvality.

#### 5.2 Experiment 2: Analýza výpočtovej náročnosti

V druhom experimente sme analyzovali vzťah medzi veľkosťou modelu a výpočtovou náročnosťou. Zistili sme, že zatiaľ čo väčšie modely produkovali kvalitnejšie výstupy, nárast výpočtovej náročnosti bol značný:

- 5M parametrov: Čas generovania = 0.8s, Využitie pamäte = 1.5GB, FLOPS = 2B
- 20M parametrov: Čas generovania = 2.5s, Využitie pamäte = 4GB, FLOPS = 8B
- 50M parametrov: Čas generovania = 5.7s, Využitie pamäte = 7GB, FLOPS = 20B

Tieto výsledky zdôrazňujú dôležitosť nájdenia optimálneho kompromisu medzi kvalitou výstupu a výpočtovou efektivitou, najmä pri práci na hardvéri s obmedzenými zdrojmi, ako bol ten náš.

#### 5.3 Experiment 3: Porovnanie s existujúcimi prístupmi

V treťom experimente sme porovnali náš optimalizovaný difúzny model so state-of-the-art GAN modelom (StyleGAN2, Karras et al., 2020) na datasete CelebA-HQ. Zistili sme, že:

- Naš difúzny model (50M parametrov,  $T = 1000$ ): FID = 22.9, IS = 7.6
- StyleGAN2 (adaptovaný na podobnú veľkosť znížením počtu kanálov a úpravou hĺbky siete): FID = 19.5, IS = 8.1

Hoci StyleGAN2 dosiahol lepšie výsledky v metrikách kvality, náš difúzny model preukázal viaceré výhody: vykazoval lepšiu stabilitu tréningu, menšiu náchylnosť na mód

kolaps, umožňoval lepšiu kontrolu nad procesom generovania a potenciálne ponúka lepšiu škálovateľnosť a adaptabilitu na rôzne úlohy. Je dôležité poznamenať, že tieto výsledky boli dosiahnuté na zmenšenej verzii datasetu a s menšími modelmi, aby sme sa prispôsobili obmedzeniam nášho hardvéru. Napriek týmto obmedzeniam sme stále schopní demonštrovať kľúčové charakteristiky a výhody difúzných modelov v porovnaní s inými prístupmi.

## 6 Analýza a diskusia

Naše experimenty poskytli niekoľko kľúčových poznatkov o efektívite difúzných modelov v multimedialnej tvorbe:

1. Vplyv parametrov: Zistili sme, že zvyšovanie počtu krokov difúzie má pozitívny vplyv na kvalitu výstupu, ale s klesajúcim prínosom. Toto zistenie naznačuje, že existuje optimálny bod, za ktorým ďalšie zvyšovanie  $T$  neprináša významné zlepšenie kvality, ale zvyšuje výpočtovú náročnosť.
2. Kompromis medzi kvalitou a efektívnosťou: Naše výsledky jasne ukazujú, že existuje trade-off medzi kvalitou generovaného obsahu a výpočtovou náročnosťou. Väčšie modely produkujú kvalitnejšie výstupy, ale za cenu značne vyššej výpočtovej náročnosti. Toto zistenie zdôrazňuje dôležitosť nájdenia optimálneho kompromisu pre konkrétne aplikácie.
3. Porovnanie s existujúcimi prístupmi: Naše porovnanie s GAN modelmi ukázalo, že difúzne modely sú konkurencieschopné v oblasti kvality generovaného obsahu. Navyše, difúzne modely vykazujú lepšiu stabilitu tréningu, menšiu náchylnosť na mód kolaps, atď., čo ich robí atraktívnymi pre širšie spektrum aplikácií.

Tieto zistenia majú významné implikácie pre dizajn a nasadenie generatívnych modelov v multimedialných aplikáciách. Naznačujú, že optimálna konfigurácia difúzneho modelu závisí od konkrétnych požiadaviek aplikácie, pričom je potrebné vyvážiť kvalitu výstupu, výpočtovú náročnosť a stabilitu tréningu.

### 6.1 Interpretácia výsledkov

Naše výsledky naznačujú, že difúzne modely ponúkajú flexibilný framework pre generovanie multimedialného obsahu. Schopnosť týchto modelov produkovať vysoko kvalitné výstupy pri rôznych konfiguráciách ich robí vhodnými pre širokú škálu aplikácií, od rýchlej produkcie náčrtov až po generovanie vysoko detailných obrazov.

Zistenie, že zvyšovanie počtu krokov difúzie má klesajúci prínos, je v súlade s teoretickými predpokladmi o konvergencii difúzneho procesu. Toto zistenie môže viesť k vývoju adaptívnych stratégií pre určenie optimálneho počtu krokov difúzie v závislosti od požadovanej kvality výstupu a dostupných výpočtových zdrojov.

Významnou technickou výzvou je aj optimalizácia nákladov na prevádzku týchto modelov. Naše experimenty ukazujú, že zvýšenie počtu krokov difúzie z  $T=500$  na  $T=1000$  prinieslo len 7.3% zlepšenie v FID skóre. Tento poznatok naznačuje, že pri implementácii difúzných modelov je dôležité hľadať optimálny pomer medzi kvalitou výstupov a výpočtovou náročnosťou, keďže výrazné navýšenie výpočtových zdrojov nemusí priniesť proporcionálne zlepšenie kvality. Pri výbere implementačného riešenia existujú rôzne možnosti. Kým veľké spoločnosti si môžu dovoliť budovať vlastnú výpočtovú infraštruktúru, pre menšie organizácie môže byť výhodnejšie využívať cloudové služby, kde platia len za reálne využité zdroje.

### 6.2 Implikácie pre dizajn generatívnych modelov

Naše zistenia majú niekoľko dôležitých implikácií pre budúci dizajn a implementáciu generatívnych modelov:

1. Adaptívne stratégie: Vývoj metód pre dynamické prispôsobovanie počtu krokov difúzie počas generovania môže viesť k optimálnemu využitiu výpočtových zdrojov.
2. Hierarchické modely: Využitie hierarchických architektúr, ktoré umožňujú generovanie obsahu v rôznych úrovniach detailu, môže poskytovať flexibilitu pri vyvažovaní kvality a výpočtovej náročnosti.
3. Hybridné prístupy: Kombinácia silných stránok difúzných modelov a iných generatívnych prístupov, ako sú GANy, môže viesť k vývoju robustnejších a efektívnejších modelov.

### 6.3 Limitácie štúdie

Napriek našim snahám o komplexnú analýzu má naša štúdia niekoľko limitácií:

1. Obmedzený rozsah datasetov: Hoci sme použili široko uznávané datasety, naše zistenia nemusia byť plne generalizovateľné na všetky typy multimedialného obsahu.
2. Hardvérové obmedzenia: Naše experimenty boli limitované dostupným výpočtovým hardvérom, čo ovplyvnilo rozsah testovaných konfigurácií.
3. Zameranie na obrazové dáta: Naša štúdia sa primárne zamerala na generovanie obrazov, a preto naše zistenia nemusia byť plne aplikovateľné na iné typy multimedialného obsahu, ako je zvuk alebo video.

## 7 Výzvy a budúce smery výskumu

Napriek značnému pokroku v oblasti difúzných modelov zostáva niekoľko kľúčových výziev, ktoré je potrebné riešiť pre ich efektívne využitie v reálnych aplikáciách. Tieto výzvy zahŕňajú technické aspekty, ako je škálovateľnosť a kontrola nad procesom generovania, ale aj širšie etické a sociálne implikácie. Zároveň sa otvárajú nové smery výskumu, ktoré môžu viesť k významným vylepšeniam týchto modelov a rozšíreniu ich aplikačného potenciálu.

Z ekonomického hľadiska predstavuje nasadenie difúzných modelov významné príležitosti aj výzvy. Naše experimenty s modelmi rôznych veľkostí (5M, 20M a 50M parametrov) ukazujú, že s rastúcou veľkosťou modelu sa zvyšuje kvalita výstupov, ale zároveň významne rastie výpočtová náročnosť. Pre organizácie je preto kľúčové nájsť rovnováhu medzi kvalitou výstupov a nákladmi na prevádzku týchto modelov. Nasadenie difúzných modelov ponúka na jednej strane potenciál pre automatizáciu a optimalizáciu kreatívnych procesov, čo môže viesť k značným úsporám v oblastiach ako marketingová produkcia, dizajn produktov či vývoj digitálneho obsahu. Na druhej strane je potrebné zvážiť vysoké počtatočné náklady na implementáciu a prevádzku týchto systémov. Pre organizácie je kľúčové nájsť rovnováhu medzi investíciami do tejto technológie a očakávaným prínosom pre ich podnikanie. Vznikajú tiež nové obchodné modely založené na poskytovaní generatívnych služieb ako služby (GaaS - Generation as a Service), ktoré sprístupňujú tieto technológie aj menším subjektom na trhu.

### 7.1 Technické výzvy

Napriek sľubným výsledkom zostávajú v oblasti difúzných modelov viaceré technické výzvy:

1. Škálovateľnosť: S rastúcou komplexnosťou generovaného obsahu rastie aj výpočtová náročnosť. Výskum efektívnych metód škálovania difúzných modelov na väčšie a komplexnejšie datasety je veľmi nutný pre ich širšie nasadenie.
2. Kontrola generovania: Zlepšenie metód pre presnú kontrolu procesu generovania, napríklad prostredníctvom pokročilých techník podmieneného generovania, zostáva dôležitou výzvou.

3. Interpretovateľnosť: Hlbšie pochopenie vnútorných mechanizmov difúzných modelov môže viesť k vývoju interpretovateľnejších modelov a lepšej kontrole nad procesom generovania.

## 7.2 Etické a sociálne implikácie

S rastúcou schopnosťou generatívnych modelov produkovať realistický obsah prichádzajú aj významné etické a sociálne otázky:

1. Dezinformácie: Schopnosť generovať realistické, ale falošné multimedialne obsahy môže prispieť k šíreniu dezinformácií. Je potrebný výskum metód pre detekciu generovaného obsahu a zvyšovanie povedomia o tejto problematike.
2. Autorské práva: Generovanie obsahu založeného na existujúcich dielach vyvoláva otázky o autorských právach a vlastníctve generovaného obsahu.
3. Bias (systematická predpojatosť alebo zaujatosť v generovanom obsahu) a reprezentácia: Je dôležité skúmať a riešiť potenciálne biasy v generovanom obsahu, ktoré môžu prameniť z tréningových dát alebo dizajnu modelu.

## 7.3 Navrhované budúce smery výskumu

Na základe našich zistení a identifikovaných výziev navrhujeme nasledujúce smery pre budúci výskum:

1. Vývoj adaptívnych difúzných procesov, ktoré dynamicky prispôbujú počet krokov a parametre modelu počas generovania.
2. Výskum hybridných architektúr, ktoré kombinujú silné stránky difúzných modelov s inými generatívnymi prístupmi.
3. Vývoj metód pre zlepšenie interpretovateľnosti a kontrolovateľnosti difúzných modelov.
4. Výskum etických implikácií a vývoj metód pre zodpovedné nasadenie generatívnych modelov v reálnych aplikáciách.

## 8 Záver

V tejto štúdií sme sa zamerali na analýzu efektivity difúzných modelov v kontexte multimedialnej tvorby, s osobitným dôrazom na vzťah medzi kvalitou generovaného obsahu a výpočtovou náročnosťou. Naše experimenty odhalili komplexnú sieť vzťahov medzi parametrami modelu, kvalitou výstupu a výpočtovou efektivitou, poskytujúc cenné poznatky pre budúce nasadenie týchto modelov v reálnych aplikáciách.

Zistili sme, že difúzne modely ponúkajú konkurencieschopnú kvalitu generovania v porovnaní s existujúcimi prístupmi, pričom vykazujú lepšiu stabilitu tréningu a robustnosť. Tieto vlastnosti robia difúzne modely sľubnou technológiou pre širokú škálu multimedialných aplikácií. Avšak, naše výsledky tiež poukazujú na potrebu ďalšieho výskumu v niekoľkých kľúčových oblastiach.

Jednou z týchto oblastí je vývoj adaptívnych stratégií pre optimalizáciu procesu difúzie. Naše experimenty naznačujú, že existuje potenciál pre dynamické prispôbovanie počtu krokov difúzie v závislosti od konkrétnej vstupnej vzorky a požadovanej kvality výstupu. Takýto prístup by mohol viesť k významným zlepšeniam v efektivite bez kompromisov v kvalite generovaného obsahu.

Ďalšou sľubnou oblasťou výskumu je integrácia difúzných modelov s inými technikami strojového učenia. Napríklad, kombinácia difúzných modelov s metódami aktívneho učenia by mohla viesť k efektívnejším stratégiám pre výber a spracovanie tréningových dát. Podobne, využitie techník transferového učenia v kontexte difúzných modelov by mohlo umožniť rýchlejšiu adaptáciu na nové domény a úlohy.



Naše výsledky tiež zdôrazňujú potrebu vývoja sofistikovanejších metrík pre hodnotenie kvality generovaného obsahu. Zatiaľ čo metriky ako FID a IS poskytujú cenné informácie, nepokrývajú všetky aspekty vnímanej kvality a užitočnosti generovaného obsahu. Vývoj nových metrík, ktoré lepšie zachytávajú sémantickú koherenciu, štylistickú konzistenciu a kontextuálnu vhodnosť generovaného obsahu, by mohol významne prispieť k ďalšiemu pokroku v oblasti generatívnych modelov.

V kontexte etických implikácií našej práce je dôležité zdôrazniť potrebu ďalšieho výskumu v oblasti interpretovateľnosti a kontrolovateľnosti difúzných modelov. Schopnosť pochopiť a vysvetliť rozhodovací proces týchto modelov je kľúčová pre ich zodpovedné nasadenie v citlivých aplikáciách. Hoci techniky ako LoRA (Low-Rank Adaptation) už umožňujú určitú mieru kontroly nad procesom generovania, stále existuje priestor pre vývoj pokročilejších metód jemnej kontroly. Ďalší výskum v tejto oblasti by mohol viesť k ešte presnejšiemu a intuitívnejšiemu ovládaniu generatívneho procesu, čím by sa otvorili nové možnosti pre tvorivé a interaktívne využitie týchto modelov. Integrácia týchto pokročilých kontrolných mechanizmov s užívateľsky priateľskými rozhraniami by mohla demokratizovať prístup k sofistikovanej multimedialnej tvorbe pre širšie spektrum užívateľov..

Z praktického hľadiska naše zistenia poukazujú na dôležitosť vyváženia výpočtovej efektivity a kvality výstupu pri nasadzovaní difúzných modelov v reálnych aplikáciách. Toto je obzvlášť relevantné v kontexte mobilných a edge computing scenárov, kde sú výpočtové zdroje často obmedzené. Výskum optimalizačných techník špecifických pre difúzne modely, ako aj vývoj hardvérových akcelerátorov prispôbených pre tieto modely, predstavujú sľubné smery pre budúci výskum a vývoj.

Nakoniec, naša práca otvára nové otázky týkajúce sa škálovateľnosti difúzných modelov. Zatiaľ čo naše experimenty sa zamerali na modely strednej veľkosti, budúci výskum by mohol skúmať správanie týchto modelov pri extrémnom škálovaní, podobne ako v prípade veľkých jazykových modelov. Takýto výskum by mohol odhaliť nové emergentné vlastnosti a schopnosti difúzných modelov, potenciálne vedúc k prelomovým objavom v oblasti generatívneho modelovania.

Záverom možno konštatovať, že difúzne modely predstavujú významný krok vpred v oblasti generatívnych modelov pre multimedialne aplikácie. S pokračujúcim výskumom a vývojom očakávame, že tieto modely budú hrať čoraz dôležitejšiu úlohu v automatizovanej tvorbe obsahu, umeleckej expresii a ďalších inovatívnych aplikáciách. Ich potenciál pre vytváranie vysoko kvalitného, diverzifikovaného a kontrolovateľného obsahu ich stavia do popredia súčasného výskumu v oblasti umelej inteligencie a multimedialnej tvorby.

Z ekonomickej perspektívy naše zistenia naznačujú, že difúzne modely predstavujú sľubnú investičnú príležitosť pre organizácie rôznych veľkostí. Zatiaľ čo implementácia vlastnej infraštruktúry môže byť nákladovo efektívna pre väčšie organizácie, rozvoj cloudových služieb a nových obchodných modelov v tejto oblasti umožňuje aj menším subjektom profitovať z týchto technológií. Kľúčovým faktorom úspechu bude schopnosť organizácií správne vyhodnotiť pomer medzi investíciami do týchto technológií a ich prínosom pre konkrétne podnikateľské prípady použitia.

Táto práca predstavuje krok k lepšiemu pochopeniu a využitiu difúzných modelov, ale zároveň odhaľuje množstvo nových výziev a príležitostí pre budúci výskum.

---

**Literatúra**

1. Briot, J. P., Hadjeres, G., & Pachet, F. (2022). Deep learning techniques for music generation. *Springer Nature*.
2. Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 8780-8794. [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf)
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>
4. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf>
5. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf>
6. Ho, J., Salimans, T., Gritsenko, A. A., Chan, W., Norouzi, M., & Fleet, D. J. (2022). Video diffusion models. *Advances in Neural Information Processing Systems*, 35, 4483-4496.
7. Karras, T., Laine, S., & Aila, T. (2018). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [https://openaccess.thecvf.com/content\\_CVPR\\_2019/papers/Karras\\_A\\_Style-Based\\_Generator\\_Architecture\\_for\\_Generative\\_Adversarial\\_Networks\\_CVPR\\_2019\\_paper.pdf](https://openaccess.thecvf.com/content_CVPR_2019/papers/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.pdf)
8. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of StyleGAN. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [https://openaccess.thecvf.com/content\\_CVPR\\_2020/papers/Karras\\_Analyzing\\_and\\_Improving\\_the\\_Image\\_Quality\\_of\\_StyleGAN\\_CVPR\\_2020\\_paper.pdf](https://openaccess.thecvf.com/content_CVPR_2020/papers/Karras_Analyzing_and_Improving_the_Image_Quality_of_StyleGAN_CVPR_2020_paper.pdf)
9. Kong, Z., Ping, W., Huang, J., Zhao, K., & Catanzaro, B. (2021). DiffWave: A versatile diffusion model for audio synthesis. *International Conference on Learning Representations*. <https://openreview.net/forum?id=a-xFK8Ymz5J>
10. Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. *Technical report, University of Toronto*. <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
11. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J. Y., & Ermon, S. (2021). SDEdit: Guided image synthesis and editing with stochastic differential equations. *International Conference on Learning Representations*.
12. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. *arXiv preprint arXiv:2204.06125*. <https://arxiv.org/abs/2204.06125>
13. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684-10695.

- 
- [https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach\\_High-Resolution\\_Image\\_Synthesis\\_With\\_Latent\\_Diffusion\\_Models\\_CVPR\\_2022\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2022/papers/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.pdf)
14. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. <https://arxiv.org/pdf/1505.04597>
  15. Saito, S., Hoshyari, S., Suzuki, T., Weng, Y., Yin, W., & Fatahalian, K. (2023). Spatially-adaptive diffusion models. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19216-19226.
  16. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf>
  17. Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. *International Conference on Machine Learning*. <http://proceedings.mlr.press/v37/sohl-dickstein15.pdf>
  18. Song, J., Meng, C., & Ermon, S. (2021). Denoising diffusion implicit models. *International Conference on Learning Representations*. <https://openreview.net/forum?id=St1giarCHLP>