

Objektivno i subjektivno ispitivanje kvalitete slika nadopunjenih s modelima dubokog učenja

Petrović, Ena

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University North / Sveučilište Sjever**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:122:239373>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-02-22**



Repository / Repozitorij:

[University North Digital Repository](#)



SVEUČILIŠTE SJEVER
SVEUČILIŠNI CENTAR VARAŽDIN



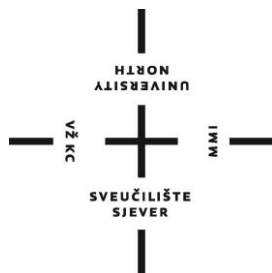
DIPLOMSKI RAD br. 137-MMD-2024

OBJEKTIVNO I SUBJEKTIVNO ISPITIVANJE
KVALITETE SLIKA NADOPUNJENIH S
MODELIMA DUBOKOG UČENJA

Ena Petrović

Varaždin, rujan 2024.

SVEUČILIŠTE SJEVER
SVEUČILIŠNI CENTAR VARAŽDIN
Studij MULTIMEDIJA



DIPLOMSKI RAD br. 137-MMD-2024

OBJEKTIVNO I SUBJEKTIVNO ISPITIVANJE
KVALITETE SLIKA NADOPUNJENIH S
MODELIMA DUBOKOG UČENJA

Student:
Ena Petrović, 0068230437

Mentor:
izv. prof. dr. sc. Emil Dumić

Varaždin, rujan 2024.

Prijava diplomskog rada

Definiranje teme diplomskog rada i povjerenstva

ODJEL	Odjel za multimediju		
STUDIJ	diplomski sveučilišni studij Multimedija		
PRISTUPNIK	Petrović Ena	IMBAG	0068230437
DATUM	26.08.2024.	KOLEGIJ	Računalni vid
NASLOV RADA	Objektivno i subjektivno ispitivanje kvalitete slika nadopunjenih s modelima dubokog učenja		
NASLOV RADA NA ENGL. JEZIKU	Objective and subjective image quality assessment inpainted with deep learning models		
MENTOR	Emil Dumić	ZVANJE	izv.prof.dr.sc.
ČLANOVI POVJERENSTVA	1. doc. art. dr. sc. Mario Periša - predsjednik 2. izv. prof. art. dr. sc. Robert Geček - član 3. izv. prof. dr. sc. Emil Dumić - mentor 4. doc. dr. sc. Andrija Bernik - zamjenski član 5.		

Zadatak diplomskog rada

BROJ	137-MMD-2024		
OPIS	U ovom radu će biti ispitana kvaliteta slika nadopunjenih s različitim modelima dubokog učenja.		
	Nadopunjavanje slike (inpainting) proizašlo je iz drevne tehnike koju su umjetnici koristili za restauraciju oštećenih slika ili fotografija s malim nedostacima kao što su ogrebotine, pukotine, prašina i mrlje, kako bi im kvaliteta bila što bliža izvornoj slici. Danas se nadopunjavanje slike naširoko primjenjuje u raznim industrijama, kao što je obrada medicinskih slika, senzorska obrada satelitskih slika, obrada slike za filmsko i umjetničko stvaralaštvo te druge primjene. Opisat će se općenito bezuvjetni i uvjetni difuzijski modeli dubokog učenja, često korišteni za generiranje novih slika. Opisat će se noviji algoritmi vezani za nadopunjavanje slike korištenim u Adobe programima, koji koriste difuzijske modele dubokog učenja: SmartBrush za nadopunjavanje tekstem i oblikom, SceneComposer za sintezu slike iz semantičkih tekstualnih opisa različitih razina te RenderDiffusion za 3D rekonstrukciju, nadopunjavanje i generiranje slika. Opisat će se Imagen Editor za lokalizirano nadopunjavanje slike iz tekstualnog opisa. Opisat će se i algoritmi za nadopunjavanje slika izvan slike (outpainting), poput FishDreamer i InOut algoritama te metode za nadopunjavanje videozapisa poput E2FGVI. Opisat će se 2 načina subjektivnog ispitivanja općenite kvalitete slika: u laboratorijskim uvjetima (ITU-R BT.500-15) ili koristeći masovno ispitivanje pomoću Interneta (crowdsourcing). Za objektivno ispitivanje kvalitete slike opisat će se različiti algoritmi koji ne koriste referentnu sliku: PIQE, NIQE te novije HyperNet, TRes i UNIQA objektivne mjere.		
	U praktičnom dijelu rada će se generirati više slika nadopunjenih s različitim algoritmima opisanim ranije, te će biti ispitana subjektivna i objektivna kvaliteta generiranih slika. U subjektivnom ispitivanju će se koristiti metoda ispitivanja putem Interneta, a objektivno ocjenjivanje će se napraviti koristeći neke od ranije navedenih objektivnih mjera kvalitete.		
ZADATAK URUČEN	02.09.2024.	POTPIS MENTORA	Emil Dumić



Sveučilište
Sjever



IZJAVA O AUTORSTVU

Završni/diplomski/specijalistički rad isključivo je autorsko djelo studenta koji je isti izradio te student odgovara za istinitost, izvornost i ispravnost teksta rada. U radu se ne smiju koristiti dijelovi tuđih radova (knjiga, članaka, doktorskih disertacija, magistarskih radova, izvora s interneta, i drugih izvora) bez navođenja izvora i autora navedenih radova. Svi dijelovi tuđih radova moraju biti pravilno navedeni i citirani. Dijelovi tuđih radova koji nisu pravilno citirani, smatraju se plagijatom, odnosno nezakonitim prisvajanjem tuđeg znanstvenog ili stručnoga rada. Sukladno navedenom studenti su dužni potpisati izjavu o autorstvu rada.

Ja, ENJA PETROVIĆ (ime i prezime) pod punom moralnom, materijalnom i kaznenom odgovornošću, izjavljujem da sam isključivi autor/ica završnog/diplomskog/specijalističkog (obrisati nepotrebno) rada pod naslovom OBJEKTIVNO I SUBJEKTIVNO ISPITIVANJE KVALITETE SLUŽBENIH VADZOVANJA (pisati naslov) te da u s navedenim radu nisu na nedozvoljeni način (bez pravilnog citiranja) korišteni dijelovi tuđih radova.

Student/ica:
(upisati ime i prezime)

Enja Petrović
(vlastoručni potpis)

Sukladno članku 58., 59. i 61. Zakona o visokom obrazovanju i znanstvenoj djelatnosti završne/diplomske/specijalističke radove sveučilišta su dužna objaviti u roku od 30 dana od dana obrane na nacionalnom repozitoriju odnosno repozitoriju visokog učilišta.

Sukladno članku 111. Zakona o autorskom pravu i srodnim pravima student se ne može protiviti da se njegov završni rad stvoren na bilo kojem studiju na visokom učilištu učini dostupnim javnosti na odgovarajućoj javnoj mrežnoj bazi sveučilišne knjižnice, knjižnice sastavnice sveučilišta, knjižnice veleučilišta ili visoke škole i/ili na javnoj mrežnoj bazi završnih radova Nacionalne i sveučilišne knjižnice, sukladno zakonu kojim se uređuje umjetnička djelatnost i visoko obrazovanje.

Predgovor

Ovim putem želim iskreno zahvaliti svom mentoru, profesoru izv. prof. dr. sc. Emilu Dumiću, na nesebičnoj podršci, stručnom vodstvu i savjetima koji su mi pomogli tijekom izrade ovog diplomskog rada. Posebno bih zahvalila svojoj obitelji i prijateljima, koji su mi kroz cijelo obrazovanje pružili podršku, motivaciju i ohrabrenje. Na kraju, želim zahvaliti Sveučilištu Sjever na svim znanjima, iskustvima i prilikama koje sam stekla tijekom proteklih pet godina. Ovo razdoblje oblikovalo je moj profesionalni i osobni razvoj, na čemu sam izuzetno zahvalna.

Sažetak

Računalni vid i napredak u dubokom učenju omogućuju razvoj modela za nadopunjavanje slika, kao što su modeli za nadopunjavanje unutar slike (engl. *inpainting*) i izvan slike (engl. *outpainting*), koji poboljšavaju vizualni sadržaj. Istraživanje ovog rada, u kojem je sudjelovalo 20 ispitanika, analiziralo je kvalitetu slika generiranih ovim modelima, uspoređujući objektivne mjere s subjektivnim ocjenama korisnika. Posebna pažnja posvećena je usporedbi alata DALL-E i Midjourney, te razlici između subjektivnog i objektivnog ocjenjivanja kvalitete slika. Zaključeno je da objektivne mjere ne mogu uvijek precizno odraziti korisničku percepciju, posebno u kategorijama autentičnosti i usklađenosti s tekstualnim opisima. Budućnost AI u generiranju slika obećava sofisticiranije i učinkovitije metode, koje će omogućiti stvaranje vizualno privlačnijih i korisnicima smislenijih slika.

Ključne riječi: umjetna inteligencija, alati, DALL-E, Midjourney, nadopunjavanje unutar slike, objektivno ocjenjivanje, kvaliteta slika...

Abstract

Computer vision and advances in deep learning enable the development of image augmentation models, such as inpainting and outpainting models, that enhance visual content. The research of this paper, in which 20 respondents participated, analyzed the quality of images generated by these models, comparing objective measures with subjective user ratings. Special attention is paid to the comparison of DALL-E and Midjourney tools, and to the difference between subjective and objective assessment of image quality. It was concluded that objective measures cannot always accurately reflect user perception, especially in the categories of authenticity and compliance with textual descriptions. The future of AI in image generation promises sophisticated and more efficient methods, which will enable the creation of more visually appealing and meaningful images for users.

Keywords: artificial intelligence, tools, DALL-E, Midjourney, inpainting, objective assessment, image quality...

Popis korištenih kratica

- HOG** engl. Histogram of Oriented Gradients
Tehnika za detekciju objekata u računalnom vidu, koja se temelji na analizi orijentacije gradijenata u slici.
- SIFT** engl. *Scale-Invariant Feature Transform*
Algoritam za detekciju i opisivanje lokalnih značajki u slikama, otporan na promjene skale i rotacije.
- CBIR** engl. *Content-Based Image Retrieval*
Tehnika pretraživanja slika u velikim bazama podataka pomoću analize sadržaja slike umjesto metapodataka.
- GLCM** engl. *Gray-Level Co-occurrence Matrix*
Matrica koja analizira teksturu slike promatranjem odnosa između piksela s različitim razinama sive u određenom prostornom odnosu.
- ID** engl. *Identification Number*
Jedinstveni broj koji se koristi za identifikaciju osobe, objekta ili entiteta u sustavima.
- R-CNN** engl. *Region-based Convolutional Neural Networks*
Metoda dubokog učenja za prepoznavanje objekata u slikama koja koristi konvolucijske neuronske mreže za generiranje prijedloga područja unutar slike i njihovu klasifikaciju.
- Brzi R-CNN** engl. *Fast Region-based Convolutional Neural Networks*
Poboljšana verzija R-CNN-a za prepoznavanje objekata, koja omogućuje bržu i učinkovitiju obradu slike, koristeći zajedničku konvolucijsku značajku za više prijedloga područja.
- YOLO** engl. *You Only Look Once*
Algoritam za detekciju objekata u slikama koji izvodi predikciju svih objekata u slici u jednom prolazu, omogućujući bržu i učinkovitiju detekciju u stvarnom vremenu.
- SOT** engl. *Single Object Tracking*
Tehnika praćenja jednog ciljanog objekta kroz niz slika ili videozapisa, često korištena u primjenama poput nadzora i računalnog vida.
- MOT** engl. *Multiple Object Tracking*
Tehnika praćenja više objekata kroz niz slika ili videozapisa, koja identificira i prati različite objekte istovremeno, često korištena u aplikacijama poput autonomnih vozila i video nadzora.
- MHT** engl. *Multiple Hypothesis Testing*
Statistička metoda koja se koristi za testiranje više hipoteza istovremeno, pri čemu se prilagođavaju kriteriji za izbjegavanje povećanja vjerojatnosti pogreške tipa I (lažno pozitivnih rezultata).
- CNN** engl. *Convolutional Neural Network*
Vrsta duboke neuronske mreže koja se koristi za analizu vizualnih podataka, posebno pogodna za prepoznavanje uzoraka u slikama putem primjene konvolucijskih slojeva.

1D engl. *One-Dimensional*
Objekti ili podaci koji imaju samo jednu dimenziju, kao što je linija ili niz brojeva, često korišteni u analizi signala i vremenskih nizova.

2D engl. *Two-Dimensional*
Objekti ili podaci koji imaju dvije dimenzije, poput širine i visine, često korišteni u računalnoj grafici, slikama, i analizama površina.

3D engl. *Three-Dimensional*
Objekti ili podaci koji imaju tri dimenzije – širinu, visinu i dubinu, često korišteni u modeliranju, računalnoj grafici, animaciji i virtualnoj stvarnosti.

MATLAB

engl. *Matrix Laboratory*

Visokorazinska programska platforma namijenjena za numeričke izračune, analizu podataka, vizualizaciju i razvoj algoritama, široko korištena u znanstvenim i inženjerskim disciplinama.

PWC-Net

engl. *Pyramid, Warping, and Cost Volume Network*

Duboka neuronska mreža dizajnirana za procjenu optičkog toka, koja koristi hijerarhijsku piramidalnu strukturu i tehnike deformacije za precizno praćenje pokreta između sličnih slika.

MCP engl. *McCulloch-Pitts Model*

Prvi matematički model umjetnog neurona, razvijen 1943. godine, koji koristi jednostavnu binarnu logiku za simulaciju načina na koji bi neuronske mreže mogle obrađivati informacije.

DBN engl. *Deep Belief Network*

Generativni grafički model sastavljen od slojeva latentnih varijabli (neuronih čvorova) koji se mogu trenirati na nesuperviziran način.

RBM engl. *Restricted Boltzmann Machines*

Stohastički generativni modeli koji koriste neuronske mreže za učenje distribucije podataka.

AE engl. *Autoencoder*

Vrsta umjetne neuronske mreže koja se koristi za učenje efikasnih kodova za nesupervizirano učenje.

VAE engl. *Variational Autoencoder*

Generativni model neuronske mreže koji koristi probabilistički pristup za kodiranje podataka.

GAN engl. *Generative Adversarial Network*

Generativni model koji se sastoji od dvije neuronske mreže - generatora i diskriminatora.

FID engl. *Fréchet Inception Distance*

Metrika koja se koristi za procjenu kvalitete generiranih slika uspoređivanjem statističkih značajki između stvarnih i generiranih slika.

- IS** engl. *Inception Score*
Metrika koja se koristi za ocjenu kvalitete generiranih slika analizom klasifikacijskih rezultata modela treniranog na skupu podataka ImageNet.
- DDPM** engl. *Denoising Diffusion Probabilistic Model*
Generativni model koji postupno dodaje šum podacima kroz više koraka, a zatim ih uči uklanjati kako bi rekonstruirao originalne podatke.
- NCSN** engl. *Noise Conditional Score Network*
Generativni model koji koristi uvjetovane bodovne funkcije za generiranje podataka, trenirajući mrežu da procijeni gradijent distribucije podataka zagađenih šumom.
- DALL-E**
Model umjetne inteligencije razvijen od strane OpenAI-a, sposoban generirati slike na temelju tekstualnih opisa.
- FoV** engl. *Field of View*
Opseg vidljivog područja koje senzor ili kamera može snimiti u određenom trenutku.
- ViT** engl. *Visual Transformer*
Model dubokog učenja koji primjenjuje mehanizam transformera na slike.
- JPEG** engl. *Joint Photographic Experts Group*
Standard za kompresiju digitalnih slika koji omogućava smanjenje veličine datoteke uz minimalni gubitak kvalitete slike.
- CFANet**
engl. *Contextual Feature Aggregation Network*
Neuronska mreža dizajnirana za poboljšanje prepoznavanja objekata u slikama, koja koristi agregaciju kontekstualnih značajki kako bi poboljšala točnost detekcije.
- ChatGPT**
engl. *Chat Generative Pre-trained Transformer*
Napredni jezični model razvijen od strane OpenAI-a, koji koristi transformere za generiranje tekstualnih odgovora na temelju unesenih podataka.
- SDXL 1.0**
Napredni model za generiranje slika temeljen na difuzijskim mrežama, koji omogućuje stvaranje visoko kvalitetnih i detaljnih slika.
- CSS** engl. *Cascading Style Sheets*
Jezik za oblikovanje koji se koristi za definiranje vizualnog prikaza web stranica napisanih u HTML-u.
- HTML** engl. *HyperText Markup Language*
Standardni jezik za izradu i strukturiranje sadržaja na web stranicama.
- URL** engl. *Uniform Resource Locator*
Jedinstvena adresa koja specificira lokaciju resursa na internetu.

Sadržaj

1.	Uvod.....	14
2.	Računalni vid	15
2.1.	Osnovne tehnike	16
2.2.	Segmentacija slike i detekcija rubova	16
2.3.	Ekstrakcija značajki.....	17
2.4.	Detekcija i praćenje objekata	18
2.5.	Rekonstrukcija slike	20
2.6.	Kalibracija kamere	21
2.7.	Optički tok.....	22
2.8.	Duboko učenje.....	23
3.	Difuzijski modeli	26
3.1.	SmartBrush.....	27
3.2.	Scene Composer	30
3.3.	RenderDiffusion	31
3.4.	Imagen Editor.....	32
3.5.	Inst-Inpaint	33
4.	Nadopunjavanje izvan slike (engl. outpainting)	35
4.1.	FishDreamer	35
4.2.	InOut.....	37
5.	Subjektivna evaluacija kvalitete slike	40
5.1.	Preporuka ITU-R BT.500-15	41
5.2.	Masovna ispitivanja (eng. <i>crowdsourcing</i>)	42
5.3.	AGIQA-3K baza podataka	44
6.	Objektivna evaluacija kvalitete slike	46
6.1.	Evaluacija kvalitete slike bez reference	47
6.2.	NIQE mjera kvalitete slike	48
6.3.	PIQUE mjera kvalitete slike.....	49
6.4.	HyperIQA mjera kvalitete slike	50
6.5.	UNIQA mjera kvalitete slike.....	51
6.6.	TOPIQ mjera kvalitete slike.....	52
7.	AI alati za generiranje slika	54
7.1.	OpenAI i ChatGPT	55
7.2.	DALL-E	55
7.3.	Midjourney	56
7.4.	Autorska prava generiranog sadržaja	57
8.	Razvoj istraživanja	59
8.1.	Priprema sadržaja za istraživanje	59
8.2.	Generiranje slika	61
8.3.	Dizajn i kreiranje web stranice	62
8.4.	Funkcionalnost web stranice	63
8.5.	Stavljanje web stranice na server	68
9.	Pregled dobivenih rezultata.....	70
9.1.	Pregled osnovnih podataka.....	70
9.2.	Usporedba subjektivnih ocjena	71

9.3. Korelacija subjektivnih i objektivnih mjera	77
10. Zaključak.....	80
11. Literatura.....	81
Popis slika	84
Popis tablica.....	87

1. Uvod

Računalni vid predstavlja područje računalne znanosti koje se bavi razumijevanjem i analizom vizualnih informacija iz digitalnih slika ili videozapisa. U kontekstu procjene kvalitete slika nadopunjenih modelima dubokog učenja, računalni vid igra ključnu ulogu u razvoju algoritama i tehnika za analizu, poboljšanje i interpretaciju kvalitete slika. Napredak u dubokom učenju omogućava razvoj sofisticiranih modela za nadopunjavanje slika, poput modela za nadopunjavanje unutar slike (engl. *inpainting*) i izvan slike (engl. *outpainting*), koji mogu popraviti ili proširiti postojeće slike koristeći naučene značajke iz velikih skupova podataka. Modeli za nadopunjavanje unutar slike koriste se za popunjavanje nedostajućih dijelova slike, dok nadopunjavanje izvan slike proširuje granice slike stvaranjem novih dijelova koji prirodno nadopunjuju originalni sadržaj. Ove tehnike omogućuju stvaranje vizualno privlačnijih i cjelovitijih slika, što je korisno u raznim primjenama. Primjene računalnog vida u kontekstu obrade slika i njihove kvalitete su raznolike i obuhvaćaju mnoge industrije i područja. Segmentacija slike i detekcija rubova dvije su ključne tehnike u računalnom vidu koje omogućuju strojevima da identificiraju različite dijelove slike i definiraju njihove granice. Detekcija rubova koristi se za pronalaženje oštrih promjena u intenzitetu piksela koje označavaju rubove objekata, dok segmentacija dijeli sliku na manje podregije koje zadovoljavaju određene uvjete, poput sličnosti intenziteta ili teksture. Ekstrakcija značajki važan je korak u analizi slike nakon segmentacije, jer omogućuje prepoznavanje i opisivanje bitnih elemenata slike. Detekcija i praćenje objekata kritični su zadaci u računalnom vidu, posebno u područjima poput autonomnih vozila i nadzornih sustava. Rekonstrukcija slike je proces koji ima za cilj poboljšanje kvalitete degradiranih slika korištenjem poznatih informacija o degradaciji. Duboko učenje revolucioniralo je računalni vid kroz primjenu neuronskih mreža koje su sposobne učiti složene obrasce iz velikih skupova podataka. Konvolucijske neuronske mreže najvažniji su tip modela u dubokom učenju za računalni vid, jer omogućuju izvanredno prepoznavanje uzoraka, objekata i scena iz slika. Difuzijski modeli nedavno su postali ključni u generiranju slika i zvuka, omogućujući visoku kvalitetu i vjernost uzoraka. Umjetna inteligencija (AI) je dio informatike koji razvija sposobnost računala za obavljanje zadataka koji zahtijevaju inteligenciju. AI sustavi mogu obrađivati velike količine informacija, komunicirati s ljudima ili uređajima, učiti iz iskustva, donositi odluke i prilagođavati se situacijama. Ipak, današnji AI sustavi su specijalizirani za određena područja, dok čovjek može istovremeno obavljati različite zadatke.

2. Računalni vid

Računalni vid predstavlja područje računalne znanosti koje se bavi razumijevanjem i analizom vizualnih informacija iz digitalnih slika ili videozapisa. U kontekstu procjene kvalitete slika nadopunjenih modelima dubokog učenja, računalni vid igra ključnu ulogu u razvoju algoritama i tehnika za analizu, poboljšanje i interpretaciju kvalitete slika. Duboko učenje, kao područje umjetne inteligencije koje se temelji na složenim neuronskim mrežama, pokazalo se izuzetno učinkovitim u obradi slika i postizanju visokih performansi u raznim zadacima računalnog vida. Ovaj napredak omogućava razvoj sofisticiranih modela za nadopunjavanje slika, kao što su modeli za nadopunjavanje unutar slike (engl. inpainting) i izvan slike (engl. outpainting), koji mogu popraviti ili proširiti postojeće slike na temelju naučenih značajki iz velikih skupova podataka. Modeli za nadopunjavanje unutar slike koriste se za popunjavanje nedostajućih dijelova slike, dok nadopunjavanje izvan slike proširuje granice slike, stvarajući nove dijelove koji prirodno nadopunjuju originalni sadržaj. Ove tehnike omogućuju stvaranje vizualno privlačnijih i cjelovitijih slika, što je korisno u raznim primjenama. Primjene računalnog vida u kontekstu obrade slika i njihove kvalitete su raznolike i obuhvaćaju mnoge industrije i područja. Na primjer, u medicinskoj dijagnostici, računalni vid se koristi za analizu medicinskih slika radi otkrivanja abnormalnosti i bolesti. Primjenom naprednih algoritama, liječnici mogu brže i preciznije dijagnosticirati stanja poput tumora, što značajno poboljšava ishode liječenja pacijenata. U filmskoj industriji, alati računalnog vida koriste se za obradu i poboljšanje kvalitete slika kako bi se postigla bolja estetika i vizualni dojam. Specijalni efekti, restauracija starih filmova i optimizacija slike za različite formate samo su neki od primjera gdje računalni vid doprinosi kvaliteti produkcije. U industriji sigurnosti, računalni vid se koristi za nadzor i analizu videozapisa radi otkrivanja sumnjivih aktivnosti. Sustavi za prepoznavanje lica, analiza ponašanja i detekcija anomalija koriste se za povećanje sigurnosti u javnim i privatnim prostorima. Kvaliteta slika nadopunjenih modelima dubokog učenja ovisi o mnogim faktorima, uključujući arhitekturu modela, kvalitetu ulaznih podataka i pravilno podešavanje hiperparametara. Razumijevanje ovih faktora i primjena naprednih tehnika računalnog vida ključni su koraci u postizanju visoke kvalitete obrade slika. Na primjer, odabir odgovarajuće arhitekture neuronske mreže može značajno utjecati na sposobnost modela da točno rekonstruira ili poboljša sliku. Kvaliteta ulaznih podataka, uključujući rezoluciju i čistoću slike, također igra ključnu ulogu, jer bolji ulazni podaci omogućuju modelima da nauče preciznije i detaljnije značajke. Kroz integraciju računalnog vida s procjenom kvalitete slika, mogu se postići značajni napredci u raznim aplikacijama, poboljšavajući tako iskustvo korisnika i rezultate u stvarnom svijetu. Na primjer, u industriji e-trgovine, poboljšane slike proizvoda mogu privući više kupaca i povećati prodaju. U autonomnim

vozilima, bolja analiza i interpretacija vizualnih informacija može poboljšati sigurnost i učinkovitost vožnje. Kontinuirani napredak u područjima računalnog vida i dubokog učenja vodi prema sofisticiranijim i učinkovitijim rješenjima koja imaju potencijal transformirati različite aspekte našeg života [1].

2.1. Osnovne tehnike

Računalni vid jedno je od najdinamičnijih područja računalne znanosti, koje se fokusira na omogućavanje računalima da razumiju i interpretiraju vizualne informacije iz svijeta oko nas. Kroz primjenu raznih algoritama i metoda, računalni vid omogućuje strojevima da procesuiraju slike i videozapise na način koji je sličan ljudskoj percepciji. Osnovne tehnike računalnog vida igraju ključnu ulogu u ovoj disciplini, pružajući temeljne alate i pristupe koji omogućuju računalima analizu i razumijevanje vizualnih podataka. Ove tehnike omogućuju prepoznavanje objekata, otkrivanje rubova, segmentaciju slika, praćenje kretanja i mnoge druge zadatke koji su neophodni za širok spektar aplikacija. Sve ove aplikacije oslanjaju se na temeljne tehnike računalnog vida, koje su evoluirale kroz godine istraživanja i razvoja. U nastavku slijedi pregled najvažnijih osnovnih tehnika računalnog vida koje omogućuju ove napredne primjene i kontinuirano potiču inovacije u ovom području [2].

2.2. Segmentacija slike i detekcija rubova

Detekcija rubova je ključna tehnika u digitalnoj obradi slika koja omogućuje identifikaciju oštih promjena u intenzitetu piksela. Te promjene često ukazuju na granice objekata unutar slike, što je od suštinske važnosti za segmentaciju i prepoznavanje oblika. Segmentacija slike može se promatrati kao proces koji dijeli prostor slike na više manjih podregija koje zadovoljavaju određene uvjete, poput povezanosti i nepodudarnosti. Svaka podregija mora biti povezana u smislu da su svi pikseli unutar regije međusobno povezani, a regije se međusobno ne smiju preklapati. Također, svaki piksel mora pripadati nekoj regiji, a susjedne regije moraju imati različite karakteristike definirane određenim kriterijima, poput intenziteta ili teksture. Metode segmentacije slika obično se temelje na dvije glavne kategorije: diskontinuitetu i sličnosti. Segmentacija bazirana na diskontinuitetu koristi promjene u intenzitetu kako bi identificirala granice između različitih regija. Glavna tehnika ovdje je segmentacija bazirana na rubovima, gdje se granice detektiraju na temelju lokalnih promjena u intenzitetu. Segmentacija bazirana na sličnosti dijeli sliku na regije koje su slične prema određenim kriterijima, poput intenziteta ili teksture. Primjerice, slika regije konstantnog intenziteta na tamnijoj pozadini može se segmentirati tako da se granica između dvije regije detektira na temelju promjena u intenzitetu. Kod složenijih teksturiranih regija,

segmentacija može biti izazovnija zbog prisutnosti mnogih malih rubova unutar regije. U tim slučajevima koriste se metode koje razlikuju teksturirane i konstantne regije na temelju statističkih mjera poput standardne devijacije intenziteta. Detekcija rubova koristi prve i druge derivacije intenziteta slike za identifikaciju rubova. Prva derivacija pruža informacije o jačini promjene intenziteta, dok druga derivacija pomaže u preciznijem lociranju rubova. Metode poput Sobelovih i Prewittovih operatora koriste različite jezgre ili kernele za računanje gradijenta slike, dok napredniji algoritmi poput Cannyjevog koriste višestupanjski pristup za detekciju rubova, uključujući glatkoću slike, računanje gradijenta, suzbijanje ne-maksimuma i histerezu praga.

Detekcija rubova je kritična komponenta u segmentaciji slika, omogućujući precizno prepoznavanje struktura i objekata unutar slike. Segmentacija slike proces je podjele slike na različite segmente ili regije koje su homogene prema određenim kriterijima, poput boje, teksture ili intenziteta. Segmentacija pomaže u izoliranju i prepoznavanju objekata unutar slike. Postoji nekoliko tehnika segmentacije, od kojih je segmentacija pomoću praga (engl. *thresholding*) jednostavna tehnika koja dijeli sliku na temelju intenziteta piksela. Globalna segmentacija pomoću praga koristi jedan prag za cijelu sliku, dok adaptivna segmentacija pomoću praga koristi različite pragove za različite dijelove slike. Metoda rastućih regija započinje s početnim pikselom i širi regiju uključivanjem susjednih piksela koji zadovoljavaju određeni kriterij sličnosti [3].

2.3. Ekstrakcija značajki

Nakon segmentacije slike, često je potrebno segmentirane piksele pretvoriti u oblik prikladan za daljnju računalnu obradu. Tipično, korak nakon segmentacije je ekstrakcija značajki, koja se sastoji od detekcije i opisa značajki. Detekcija značajki odnosi se na pronalaženje značajki u slici, regiji ili granici, dok opis značajki dodjeljuje kvantitativne atribute detektiranim značajkama. Na primjer, možemo detektirati uglove u granici regije i opisati te uglove prema njihovoj orijentaciji i lokaciji, što su oba kvantitativna atributa [4]. Metode za obradu značajki podijeljene su u tri glavne kategorije: one koje se primjenjuju na granice, regije ili cijele slike. Neke značajke su primjenjive na više od jedne kategorije. Deskriptori značajki trebaju biti što je moguće manje osjetljivi na varijacije parametara poput skale, translacije, rotacije, osvjetljenja i pogleda. Deskriptori koji se obrađuju mogu biti neosjetljivi na te varijacije ili se mogu normirati kako bi kompenzirali te promjene. Iako ne postoji univerzalno prihvaćena formalna definicija što čini značajku slike, intuitivno smatramo značajku kao prepoznatljivi atribut ili opis "nečega" što želimo označiti ili razlikovati. "Nešto" od interesa može se odnositi na pojedinačne objekte slike, ili čak na cijele slike ili baze slika. Dakle, značajke se smatraju kao atributi koji će nam pomoći u dodjeljivanju jedinstvenih oznaka objektima na slici ili, općenitije, biti korisne u razlikovanju cijelih slika ili obitelji slika. Da bi proces ekstrakcije bio koristan, mora obuhvaćati oba aspekta.

Ključne metode za ekstrakciju značajki uključuju Histogram orijentiranih gradijenata (HOG), koji koristi histogram distribucije orijentacija gradijenata unutar lokalnih područja slike. Ova tehnika je korisna za detekciju objekata jer bilježi strukturu rubova ili gradijenata koja je karakteristična za lokalni oblik, čak i bez preciznog znanja o odgovarajućim pozicijama gradijenata ili rubova. Transformacija značajki neovisna o skali (engl. *Scale-Invariant Feature Transform*, SIFT) je metoda za detekciju i opis lokalnih značajki u slikama koje su neovisne na skaliranje, rotaciju i promjene osvjetljenja. Algoritam identificira ključne točke na slici i računa deskriptore na temelju lokalnih orijentacija gradijenata oko svake ključne točke. Ova robusnost čini SIFT vrlo korisnim za prepoznavanje objekata, spajanje slika i druge zadatke gdje se mogu mijenjati uvjeti snimanja ili osvjetljenja. Valićni momenti ili Wavelet momenti (engl. *wavelet moments*) i nepromjenjivi momenti (engl. *moment invariants*) slike su metode koje se koriste za ekstrakciju značajki teksture, analizirajući uzorke teksture u slici. Valićni momenti (ili Wavelet momenti) mogu zabilježiti i prostorne i frekvencijske informacije, čineći ih učinkovitima za analizu i klasifikaciju tekstura. Nepromjenjivi momenti slike su značajke koje ostaju nepromijenjene pod određenim transformacijama, poput translacije, skaliranja i rotacije. Pretraživanje slike prema sadržaju (engl. *Content-Based Image Retrieval*, CBIR) tehnike često se oslanjaju na kombinaciju boje, teksture i oblika kako bi pretražile slike iz velike baze podataka na temelju sadržaja upitne slike. Ove tehnike mogu koristiti razne metode ekstrakcije značajki, uključujući histogram boja, matrica supojavljivanja (engl. *gray-level co-occurrence matrix*, GLCM) i deskriptore oblika [3].

2.4. Detekcija i praćenje objekata

Praćenje objekata u računalnom vidu je proces kojim algoritam detektira i predviđa položaje cilja u nizu slika ili videozapisa. To je ključan zadatak u računalnom vidu, te uključuje više potprocesa. Prije nego što se može pratiti kretanje objekta, potrebno ga je prvo detektirati. Detekcija objekata locira prisutnost objekata u slici pomoću graničnih okvira (engl. *bounding boxova*) (lokalizacija) i identificira tipove objekata (klasifikacija). Nakon detekcije, svakom objektu se dodjeljuje jedinstveni identifikacijski broj (engl. *identification number*, ID) za praćenje kroz niz slika ili videozapisa. Konačno, algoritam procjenjuje položaje objekata u preostalim okvirima kako bi dobio njihove putanje. Popularni algoritmi za detekciju objekata uključuju R-CNN (engl. *Region-based Convolutional Neural Networks*), Brzi R-CNN (engl. *Fast R-CNN*), Brži R-CNN (engl. *Faster R-CNN*), maskirani R-CNN (engl. *Mask R-CNN*) i YOLO (engl. *You Only Look Once*). Postoje dvije glavne vrste praćenja objekata: praćenje jednog objekta (engl. *single object tracking*, SOT) i praćenje više objekata (engl. *multiple object tracking*, MOT). SOT prati jedan objekt kroz kontinuirane video okvire, poput praćenja lopte tijekom košarkaške utakmice, gdje se početni položaj cilja u prvom okviru daje algoritmu za praćenje (engl. *tracker*),

koji zatim procjenjuje položaj objekta u preostalim okvirima. MOT, s druge strane, prati više objekata istovremeno, poput praćenja igrača tijekom košarkaške utakmice, i može pratiti više objekata iz različitih klasa, kao što su pješaci, vozila i prometni znakovi u autonomnim vozilima. Praćenje objekata omogućuje kontinuirano promatranje ciljnog objekta tijekom vremena, pružajući bogatije informacije od same detekcije objekata. Primjerice, detekcija objekata može odgovoriti na pitanja poput broja ljudi u trgovini u određeno vrijeme, ali ne može pružiti informacije o njihovom kretanju prije i poslije tog trenutka. Praćenje objekata, s druge strane, omogućuje povezivanje istog objekta u različitim okvirima, pružajući uvid u njegovu putanju i ponašanje. Praćenje objekata može biti složeno zbog brojnih faktora kao što su brzi pokreti objekata, zaklanjanje, šum (engl. *noise*), rotacija, promjene skale i pokretne kamere. Ovi izazovi su posebno izraženi kod praćenja malih objekata na velikim udaljenostima. Mali objekti, poput udaljenih padobranaca ili lopti, često su teško uočljivi zbog svoje male veličine i mogu se zamijeniti za šum, što negativno utječe na točnost praćenja. Mali objekti prikazani su na slici 2.1. Različite tehnike i metode koriste se za učinkovitu detekciju i praćenje objekata u slikama i video zapisima. Filter-based metode koriste Kalmanov i čestični filtar (engl. *particle filter*) za praćenje objekata, omogućavajući upravljanje zaklanjanjem i pozadinskim smetnjama te poboljšanje vizualne kvalitete malih objekata kroz super-rezoluciju. Na primjer, adaptivni čestični filtar može značajno poboljšati točnost praćenja objekata u dinamičnim okruženjima. Metode temeljene na pretraživanju (engl. *search-based*) uključuju temeljitu pretragu kako bi pronašle najbolje tragove objekata. Jedan od primjera je Višestruko testiranje hipoteza (engl. *multiple hypothesis testing*, MHT) koji koristi intenzitet objekta za praćenje i prepoznavanje njegovog kretanja kroz sekvencu slika. Ove metode su korisne u scenarijima gdje je potrebno procijeniti više mogućih puteva kretanja objekta. Metode temeljene na detekciji pozadine se oslanjaju na razlike između okvira i oduzimanje pozadine kako bi detektirali i pratili pokretne objekte. Adaptivno oduzimanje pozadine, na primjer, omogućava učinkovitu detekciju i praćenje objekata u videosekvencama, prilagođavajući se promjenama u osvjetljenju i pozadinskim uvjetima. Klasične metode računalnog vida koriste matematičke modele i heuristike za detekciju i praćenje objekata. Ovi pristupi često uključuju algoritme temeljene na frekvencijskoj i prostornoj domeni, što je posebno korisno za praćenje malih objekata u infracrvenim slikama. Metode temeljene na dubokom učenju koriste duboke neuronske mreže za detekciju i praćenje malih objekata. Primjeri uključuju korištenje konvolucijskih neuronskih mreža engl. (*convolutional neural network*, CNN) za učenje značajki i klasifikatora, kao i YOLOv3 algoritam koji omogućava detekciju i praćenje objekata u stvarnom vremenu. Transformer neuronske mreže također je postao popularan za praćenje malih objekata iz zraka, zahvaljujući sposobnosti da obrađuje složene vizualne podatke [5].

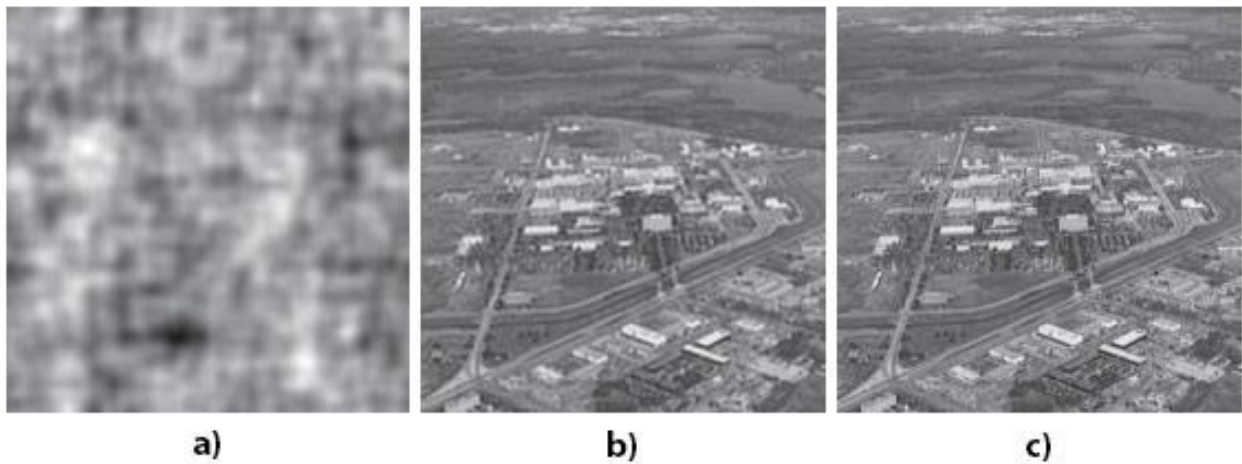


Slika 2.1 Reprezentativni primjerci malih objekata [5].

2.5. Rekonstrukcija slike

Rekonstrukcija slike je proces kojim se poboljšava degradirana slika korištenjem unaprijed poznatih informacija o degradaciji. Cilj rekonstrukcije je vratiti izvornu sliku koja je degradirana, korištenjem modela degradacije i primjenom obrnutog procesa kako bi se dobila što sličnija slika originalu. Jedan od osnovnih modela za rekonstrukciju je linearni, prostorno invariantni model, koji predstavlja degradaciju slike kao operator koji djeluje na ulaznu sliku uz dodatak šuma, te proizvodi degradiranu sliku. Šum u digitalnim slikama može nastati tijekom prijenosa slike, a može biti izazvan varijacijama u osvjetljenju, temperaturi senzora ili atmosferskim poremećajima. U procesu rekonstrukcije slike, šum se često modelira kao aditivni šum. Metode prostornog filtriranja koriste se za procjenu izvorne slike, uključujući filtere kao što su aritmetički, geometrijski, harmonijski i kontraharmonijski usrednjavajući filtar. Za složenije situacije, kao što su degradacije uzrokovane periodičkim šumom, koriste se metode filtriranja u frekvencijskoj domeni, poput notch filtera koji uklanjaju frekvencijske komponente šuma iz Fourierovog spektra slike. Adaptivni filteri prilagođavaju svoje ponašanje na temelju lokalnih statističkih karakteristika slike, pružajući bolje rezultate jer se prilagođavaju varijacijama unutar slike. Kada potrebne veličine za rekonstrukciju nisu poznate, koristi se aproksimacija s konstantom K dodanom svim članovima funkcije degradacije. Wienerov filtar koristi statistički pristup za minimiziranje srednje kvadratne pogreške, što rezultira kvalitetnijom slikom bližom originalu uz efikasno uklanjanje šuma. Filtriranje uz minimizaciju srednje kvadratne pogreške zahtijeva samo osnovne statističke informacije o šumu, optimizirajući rezultat za svaku pojedinačnu sliku. Iterativni algoritam za procjenu optimalnog filtra prilagođava karakteristike filtra kroz ponavljanje koraka izračunavanja ostatka između degradirane i procijenjene slike. Rekonstrukcija slike iz serije projekcija, poput CT-a, koristi Radonovu transformaciju. Proces uključuje dobivanje 1D projekcija 2D objekta iz različitih kutova i kombiniranje tih projekcija za formiranje 2D slike putem povratne projekcije. Filtriranje projekcija prije povratne projekcije smanjuje zamućenje i artefakte, poboljšavajući kvalitetu konačne slike. Rekonstrukcija slike kombinira statističke metode i transformacije kako bi se poboljšala kvaliteta slike, smanjujući zamućenje i artefakte te osiguravajući preciznije i

jasnije slike. U metode rekonstrukcije slike spada i tehnika koja popunjava nedostajuće dijelove slike (engl. *inpainting*) na temelju okolnih piksela koja će biti objašnjena u nastavku. Na slici ispod prikazana je usporedba inverznog i Wienerovog filtriranja [3].



Slika 2.2 Usporedba inverznog i Wienerovog filtriranja. (a) Rezultat inverznog filtriranja slike na slici, (b) Rezultat inverznog filtriranja s radijalnim ograničenjem, (c) Rezultat korištenjem Wienerovog filtra [3].

2.6. Kalibracija kamere

Kalibracija kamere je proces određivanja intrinzičnih parametara kamere, uključujući žarišnu duljinu i glavnu točku i posmak te ekstrinzičnih parametara koji predstavljaju rotaciju i translaciju kamere u odnosu na svijet. Ovaj proces je ključan za mnoge primjene u računalnom vidu, posebno kada je potrebna metrička informacija o sceni. Za kalibraciju kamere često se koristi model kamere s rupicom, a za kamere s promjenjivim zumom i fokusom kalibracija se mora prilagoditi promjenama žarišne duljine. Kalibracija se obično izvodi pomoću uzorka za kalibraciju s poznatom geometrijom, često šahovske ploče. Proces uključuje nekoliko koraka. Prvo se postavlja uzorak za kalibraciju u različite položaje ispred kamere kako bi se snimile slike iz različitih kutova. Zatim se detektiraju točke na uzorku i njihove 2D koordinate uspoređuju s 3D koordinatama u stvarnom svijetu. Nakon toga se izračunavaju intrinzični i ekstrinzični parametri kamere, koristeći linearne ili nelinearne metode. Za kalibraciju kamere dostupni su razni alati kao što su MATLAB-ov Toolbox i OpenCV biblioteka. Ovi alati omogućuju automatsko detektiranje uzorka šahovske ploče i izračunavanje potrebnih parametara [6]. Metode kalibracije uključuju ručno postavljanje referentnog niza markera na kameru kako bi se uspostavila geometrijska veza između slike kamere i referentnog niza. Nakon što je to postignuto, virtualne informacije iz skupa podataka mogu se uskladiti sa slikom kamere. Jedan od praktičnih primjera kalibracije kamere uključuje primjenu ARToolkit-a, koji koristi specifične uzorke u slici za procjenu položaja kamere. Kamera se

postavlja iznad uzorka, a alat određuje položaj kamere u odnosu na marker. Ova metoda omogućava preciznu kalibraciju koja se mora ponoviti samo ako se pomakne referentni niz kamere. Kalibracija kamere je ključan korak u mnogim tehnikama 3D rekonstrukcije. Kalibracija je neophodna za precizno mapiranje 3D točaka iz stvarnog svijeta na 2D slike koje kamera snima. Jednu od prvih praktičnih metoda kalibracije predložio je Tsai 1987. godine, koja koristi jednu sliku kalibracijskog uzorka, dok je kasnija metoda koju je predložio Zhang 2000. godine zahtijevala više promatranja uzorka iz različitih kutova. Ove metode omogućuju točniju i fleksibilniju kalibraciju. Kalibracijski proces često uključuje snimanje uzorka, poput šahovske ploče, iz više kutova. Detekcija točaka na uzorku i njihovo mapiranje između 2D koordinata na slici i 3D koordinata u stvarnom svijetu omogućuje izračun potrebnih parametara kamere. Fleksibilnost i stabilnost kalibracije omogućuju njenu primjenu u različitim sustavima vizije, uključujući perspektivne i kamere "riblje oko" (engl. *fish-eye*), te u složenim sustavima s više kamera. Takva preciznost u kalibraciji je ključna za aplikacije poput proširene stvarnosti i kirurške navigacije, gdje je važno precizno integrirati virtualne informacije sa stvarnim slikama. [7]

2.7. Optički tok

Optički tok metoda je procjene gibanja u slikama, sa fokusom na procjenu gibanja svakog pojedinog piksela. Uključuje smanjenje razlike u svjetlini ili boji između odgovarajućih piksela u sekvencama slika. Klasični pristupi ovom problemu uključuju procjenu gibanja u malim, preklapajućim dijelovima slike ili dodavanje uvjeta glatkoće pomoću matematičkih metoda kako bi se postigao najbolji ukupni rezultat. Metoda zasnovana na regijama koristi Taylorov razvoj funkcije pomaknute slike za precizne procjene manje od piksela. Horn i Schunck razvili su algoritam koji minimizira razliku svjetline za sve vektore optičkog toka tj. računa globalnu procjenu toka, dodajući uvjete glatkoće kako bi se ograničio problem. Ova tehnika je posebno prikladna za male pomake između uzastopnih slika. Alternativno, moguće je kombinirati lokalne i globalne procjene toka u jedinstveni algoritam korištenjem agregiranih Hesseovih matrica svjetline. Kada je gibanje uzrokovano kamerom u statičkom okruženju, problem se može preformulirati kao procjena dubine po pikselu zajedno s parametrima globalnog gibanja kamere. Mogu se također procijeniti afini modeli gibanja po segmentima slike zajedno s korekcijama po pikselu. Budući da svjetlina slike nije uvijek prikladna za mjerenje konzistentnosti izgleda, koriste se i drugi metrički pristupi poput gradijenata slike ili faza filtara upravljivih smjerova, omogućujući procjenu lokalnih faza neovisno o promjenama osvjetljenja i kontrasta. Optički tok se procjenjuje globalnom optimizacijom, koja općenito daje bolje rezultate kod diskontinuiteta gibanja. Koriste se robusne metrike poput L1 norme kako bi se osigurala točnost procjene. Anizotropne priorne glatkoće dodatno poboljšavaju rezultate. Zbog velike dvodimenzionalne

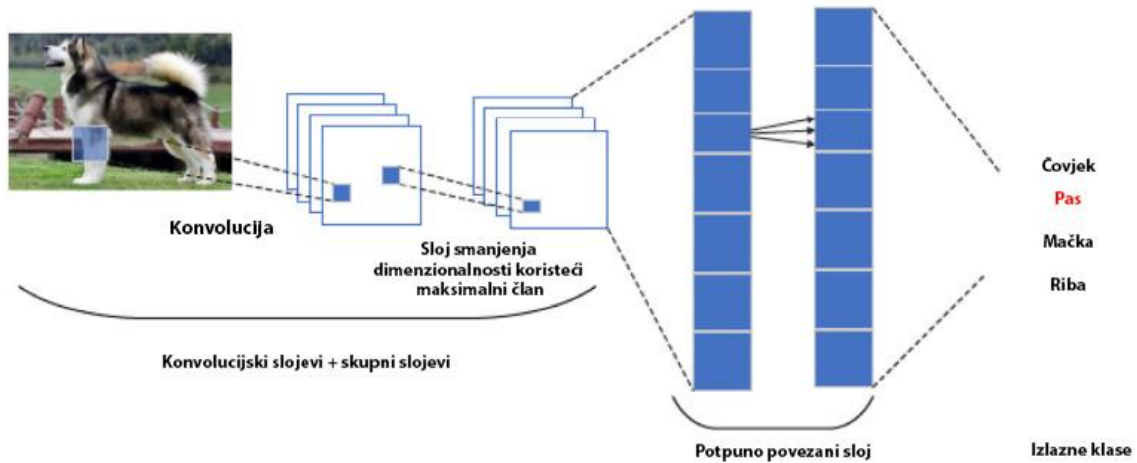
pretrage prostora, većina algoritama koristi varijante gradijentnog spusta i metode piramide od grube prema finoj rezoluciji. Nedavna dostignuća uključuju primjenu dubokih neuronskih mreža koje su postale ključne za algoritme visokih performansi optičkog toka. Prvi sustav koji je koristio potpuno učenje putem dubokih neuronskih mreža bio je FlowNet, koji je treniran na sintetičkim podacima. Kasnije verzije, kao što je PWC-Net, koriste piramide značajki zajedno s preklapanjem i rafiniranjem toka. Moderna istraživanja također uključuju modeliranje neizvjesnosti procjene toka zbog homogenih i zaklonjenih područja. Optički tok se primjenjuje u raznim zadacima kao što su uklanjanje zamućenja od pomaka objekata (engl. *motion blur*) zbog preduge ekspozicije, procjena gibanja u više kadrova, te uklanjanje šuma iz videozapisa. Algoritmi za uklanjanje zamućenja od pomaka koriste precizne procjene toka po pikselu kako bi kompenzirali izobličenja uzrokovana različitim vremenima ekspozicije za različite redove piksela [2]. Bitno je istaknuti i Lucas-Kanade metodu za procjenu optičkog toka. Lucas-Kanade metoda je ključni algoritam za procjenu optičkog toka, koji se temelji na lokalnim aproksimacijama i minimizaciji kvadratne pogreške između susjednih regija slika u nizu. Ova metoda, koju su razvili Bruce D. Lucas i Takeo Kanade 1981. godine, koristi se za praćenje gibanja objekata unutar sekvence slika. Lucas-Kanade metoda polazi od pretpostavke da je optički tok unutar malog susjedstva piksela konstantan. Algoritam koristi diferencijalne slike kako bi procijenio gibanje između dva uzastopna okvira. To postiže minimizacijom kvadratne pogreške u promjenama intenziteta piksela, što rezultira sustavom linearnih jednadžbi koje se rješavaju najmanjim kvadratima. Koraci algoritma su: kalkulacija gradijenata, zatim formiranje sustava jednadžbi i rješavanje sustava. Kalkulacijom gradijenata se prvo izračunavaju prostorni i vremenski gradijenti slike. Formiranje sustava jednadžbi opisuje gibanje svakog piksela unutar malog susjedstva. Sustav jednadžbi se rješava najmanjim kvadratima kako bi se procijenile komponente gibanja za svaki piksel. Primjena povećanjem razlučivosti za složenije scenarije s velikim pokretima, koristi se piramidalna implementacija Lucas-Kanade metode. Slike se iterativno smanjuju u razlučivosti (piramidalni pristup), a procjene gibanja iz grubljih razlučivosti koriste se kao početne vrijednosti za finije razlučivosti. Ovaj pristup omogućuje precizniju procjenu velikih gibanja. Lucas-Kanade metoda je efikasna za male, lokalne pokrete i dobro funkcionira u prisutnosti malih promjena intenziteta. Međutim, njezina preciznost može biti ograničena u prisutnosti velikih gibanja ili jakih promjena intenziteta unutar lokalnog susjedstva [8].

2.8. Duboko učenje

Duboko učenje, jedan od najvažnijih napredaka u računalnom vidu, značajno je unaprijedilo sposobnost modela da prepoznaju i razumiju složene obrasce u velikim skupovima podataka. Ovaj pristup omogućava računalnim modelima učenje i predstavljanje podataka na više razina

apstrakcije, imitirajući način na koji ljudski mozak opaža i razumije informacije. Duboko učenje uključuje različite metode poput neuronskih mreža, hijerarhijskih probablističkih modela te različitih algoritama za učenje značajki, bilo nadziranih ili nenadziranih. Početna ideja o stvaranju sustava koji simulira ljudski mozak dovela je do razvoja neuronskih mreža. Jedan od ključnih doprinosa bio je model McCulloch-Pitts (MCP) neurona koji su razvili McCulloch i Pitts 1943. godine. Taj je model postavio temelje za razvoj umjetnih neuronskih mreža. Veliki napredak u dubokom učenju dogodio se 2006. godine kada su Hinton i suradnici predstavili duboke mreže vjerovanja (engl. *Deep Belief Network*, DBN), koristeći višeslojne ograničene Boltzmannove strojeve (engl. *Restricted Boltzmann Machines*, RBM) za nenadzirano učenje. Pojava velikih, visokokvalitetnih, javno dostupnih označenih skupova podataka, zajedno s napretkom u paralelnom računalstvu putem GPU-a, omogućila je značajno ubrzanje procesa treniranja dubokih modela. To je, uz dodatne čimbenike kao što su nove tehnike regularizacije npr. izbacivanje nekih težina (engl. *dropout*), usrednjavanje slojeva (engl. *batch normalization*) i pojavu moćnih okvira poput TensorFlowa i PyTorch, dodatno unaprijedilo područje dubokog učenja. Duboko učenje donijelo je velike uspjehe u raznim problemima računalnog vida kao što su prepoznavanje objekata, praćenje pokreta, prepoznavanje akcija i aktivnosti, te procjena ljudske poze. Među najvažnijim tipovima modela dubokog učenja u ovom kontekstu su konvolucijske neuronske mreže (CNN), duboki Boltzmannovi strojevi (engl. *deep Boltzmann machine*, DBM) i slojeviti autoenkoderi (engl. *autoencoder*, AE). CNN-ovi, inspirirani strukturom vizualnog sustava koji koriste konvolucijske i potpuno povezane slojeve za ekstrakciju i obradu značajki iz slika, omogućavajući vrlo precizno prepoznavanje uzoraka i objekata. Duboki Boltzmannovi strojevi i duboke mreže vjerovanja koriste ograničene Boltzmannove strojeve kao temeljni modul za učenje, omogućavajući stvaranje hijerarhijskih reprezentacija podataka. Slojeviti autoenkoderi, s druge strane, koriste neuronske mreže za kodiranje i dekodiranje podataka, čime se postiže smanjenje dimenzionalnosti i uklanjanje šuma iz podataka. Metode koje koriste duboko učenje su također primijenjene u videoobradi, omogućujući poboljšanje kvalitete videosadržaja, smanjenje šuma i interpolaciju slika. Kombinacija tehnika dubokog učenja s naprednim algoritmima optimizacije i velikim skupovima podataka rezultirala je značajnim napretkom u performansama i točnosti računalnog vida. Razvoj generativnih modela koristeći neuronske mreže kao što su varijacijski autoenkoderi (engl. *variational autoencoder*, VAE) i generativne suparničke mreže (engl. *generative adversarial network*, GAN), također su značajno pridonijeli napretku u računalnom vidu. Kod VAE, za razliku od AE, koder preslikava svaku točku ulaza u distribuciju unutar latentnog prostora, umjesto u jednu točku u tom prostoru. Time se može koristiti i kao generativni model, za generiranje novih izlaza iz uzoraka latentnog prostora. GAN mreže se mogu koristiti za generiranje realistične slike, koriste se u aplikacijama kao što su stvaranje slika, restauracija i

stilizacija. Svaka nova generacija mreža donosi poboljšanja u točnosti i učinkovitosti. AlexNet je prvi pokazao potencijal dubokog učenja, VGGNet je povećao dubinu mreže, GoogleNet i Inception moduli su balansirali dubinu i širinu, ResNet je uveo rezidualne blokove, DenseNet je povezoao sve slojeve međusobno, MobileNets su omogućile efikasnost za mobilne uređaje, dok su EfficientNet i RegNet dodatno optimizirali performanse. CNN arhitekturu za klasifikaciju slika možemo vidjeti na Slici 2.3 [1].



Slika 2.3 CNN arhitektura za klasifikaciju slika [1].

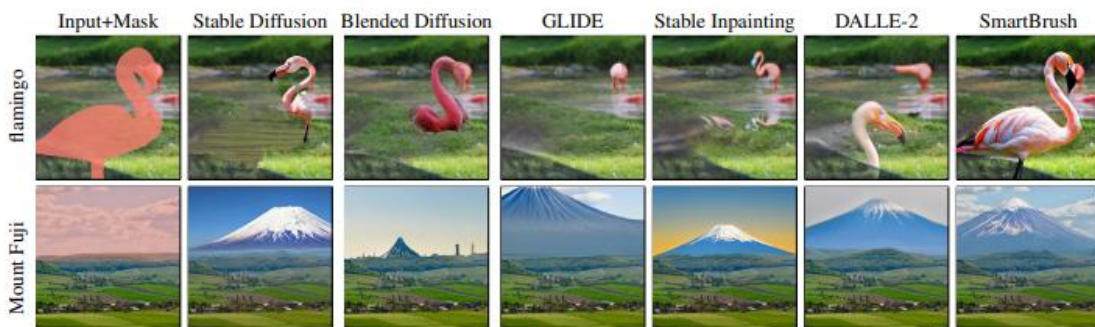
3. Difuzijski modeli

Difuzijski modeli su tehnike koje se koriste za obradu slika, a cilj im je poboljšati kvalitetu slike, ukloniti šum, ili čak nadopuniti nedostajuće dijelove slike. Vođenje klasifikatora metoda je kojom se postiže ravnoteža između pokrivenosti modela i vjernosti uzorcima u uvjetnim modelima difuzije nakon obuke, slično kao uzorkovanje s niskom temperaturom ili skraćivanje u drugim vrstama generativnih modela. Ova tehnika kombinira ocjenu difuzijskog modela s gradijentom klasifikatora slike, što zahtijeva odvojenu obuku klasifikatora slike od difuzijskog modela. Postavlja se pitanje može li se vođenje obavljati bez klasifikatora. Pokazalo se da je vođenje moguće izvesti pomoću čistog generativnog modela, bez potrebe za takvim klasifikatorom. U tom pristupu zajednički se obučavaju uvjetni i neuvjetni model difuzije, a rezultirajuće procjene ocjene kombiniraju kako bi se postigla ravnoteža između kvalitete i raznolikosti uzoraka, slično onome što se postiže korištenjem vođenja klasifikatora. Nedavno su difuzijski modeli postali izražajan i fleksibilan skup generativnih modela, pružajući konkurentnu kvalitetu uzoraka i vjerojatnost rezultata u zadacima sinteze slika i zvuka. Vođenje bez klasifikatora miješa procjene ocjena uvjetnog i ne uvjetnog modela difuzije umjesto korištenja gradijenta klasifikatora. Rezultati vođenja bez klasifikatora pokazuju da čisti generativni difuzijski modeli mogu stvarati uzorke visoke vjernosti, slične onima mogućima kod drugih vrsta generativnih modela [10]. U proteklih nekoliko godina, generativni modeli stekli su sposobnost generiranja prirodnog jezika sličnog ljudskom, beskrajno kvalitetnih sintetskih slika te izrazito raznolikog ljudskog govora i glazbe. Ovi modeli mogu se koristiti na različite načine, poput generiranja slika iz tekstualnih upita. I dok su ovi modeli već sposobni proizvoditi realistične slike i zvuk, još uvijek postoji mnogo prostora za poboljšanja, te bi bolji generativni modeli mogli imati širok utjecaj na grafički dizajn, igre, glazbenu produkciju i brojna druga područja. Generativna suparnička mreža je vrsta neuronske mreže koja se koristi za generiranje novih podataka, poput slika, zvuka ili teksta. GAN se sastoji od dva glavna dijela: generatora i diskriminatora, koji su obično neuronske mreže. GAN-ovi su do nedavno držali vodeću poziciju u većini zadataka generiranja slika mjerenih metrikama kvalitete uzoraka poput FID (Fréchet Inception Distance), IS (Inception Score) i standardne mjere (preciznost, odziv, F1 mjera). Osim toga, GAN-ovi su često teški za treniranje, s tendencijom kolapsa bez pažljivo odabranih hiperparametara i regularizatora. Kao rezultat toga, mnogo se truda ulaže kako bi se postigla kvaliteta uzoraka slična GAN-ovima s difuzijskim modelima. Difuzijski modeli, predloženi 2015 [29], imaju za cilj naučiti proces difuzije koji prvo generira distribuciju vjerojatnosti za dati skup podataka, iz koje se zatim mogu uzorkovati novi elementi reverznim postupkom. Probabilistički model uklanjanja šuma difuzijom (engl. denoising diffusion probabilistic model, DDPM [30]) uči reverzni postupak u kojem se u koracima uzorak vraća iz

Gaussovog šuma u novi element (npr. novo generiranu sliku). Svaki korak u procesu difuzije prema naprijed približno se poništi reverznim procesom treniranog difuzijskog modela. Generativni model temeljen na bodovnoj funkciji (engl. score function) još je jedan način opisivanja difuzijskih modela (šumom uvjetovana bodovna mreža, engl. noise conditional score network, NCSN [31]). Difuzijski modeli nedavno su pokazali sposobnost generiranja visokokvalitetnih slika nudeći poželjna svojstva poput dobre reprezentacije distribucije i jednostavne proširivosti. Pretpostavlja se da jaz između difuzijskih modela i GAN-ova proizlazio barem iz dva faktora: prvo, arhitekture modela korištene u nedavnoj literaturi o GAN-ovima temeljito su bile istražene i usavršene; drugo, GAN-ovi su sposobni zamijeniti raznolikost za vjernost, proizvodeći uzorke visoke kvalitete, ali ne pokrivajući cijelu distribuciju. Cilj je bio donijeti te prednosti difuzijskim modelima, prvo poboljšavanjem arhitekture modela, a zatim smišljanjem sheme za zamjenu raznolikosti za vjernost. S ovim poboljšanjima postigla se nova razina, nadmašujući GAN-ove na nekoliko različitih mjera i skupova podataka [11].

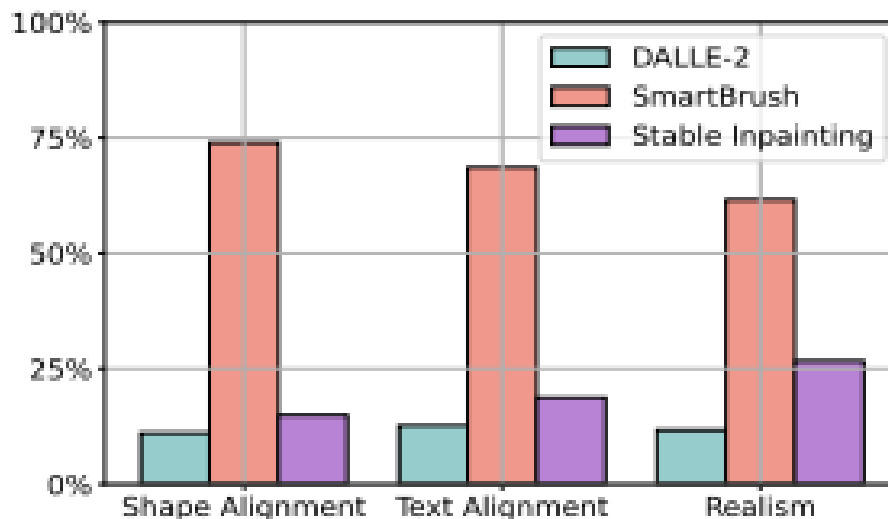
3.1. SmartBrush

Generičko nadopunjavanje slika ima cilj popuniti nedostajuće dijelove slika koristeći okolne informacije, često generirajući ograničen novi sadržaj. Nasuprot tome, višemodalno nadopunjavanje pruža fleksibilniju kontrolu nad sadržajem, omogućujući korištenje dodatnih informacija poput tekstualnih opisa i maski kako bi se preciznije oblikovao popunjeni objekt. Novi model temeljen na difuziji nazvan SmartBrush služi za nadopunjavanje nedostajućeg područja objektom koristeći i tekstualno i oblikovno vođenje [12]. Model uključuje preciznu kontrolu uz pomoć oba vodiča te uvodi inovativnu strategiju obuke i uzorkovanja kako bi se očuvala pozadina. Kroz opsežna ispitivanja, model nadmašuje sve usporedne metrike u pogledu vizualne kvalitete, kontrole maske i očuvanja pozadine, pokazujući bolje rezultate od konkurencije.



Slika 3.1 Generiranje rezultata visokokvalitetnog nadopunjavanja objekata [12].

SmartBrush može nadopuniti ne samo objekte već i općenite scene poput zalazaka sunca prateći tekstualne smjernice i oblik. Za nadopunjavanje objekata prema istraživanju u ovom članku, uzeta su dva uobičajena slučaja: 1) precizne maske objekta i 2) maske okvira. Očekuje se da će prvi generirani objekt slijediti oblik dane maske, dok drugi ne ograničava oblik generiranih objekata sve dok su unutar okvira. Kao referentna točka, model Stable Inpainting ostvaruje niže CLIP ocjene u usporedbi s predloženim SmartBrushom, ukazujući na to da je nasumično maskiranje neprikladna strategija obuke za nadopunjavanje vođeno tekstom. Model Blended Diffusion postiže relativno visoku CLIP ocjenu, ali značajno zaostaje u FID-u, s obzirom da se model CLIP usredotočuje na globalni sadržaj umjesto na lokalne objekte. Nasuprot tome, SmartBrush ostvaruje najbolje rezultate u oba zadatka, što potvrđuje učinkovitost predložene strategije treninga s tekstom i smjernicama za oblik.

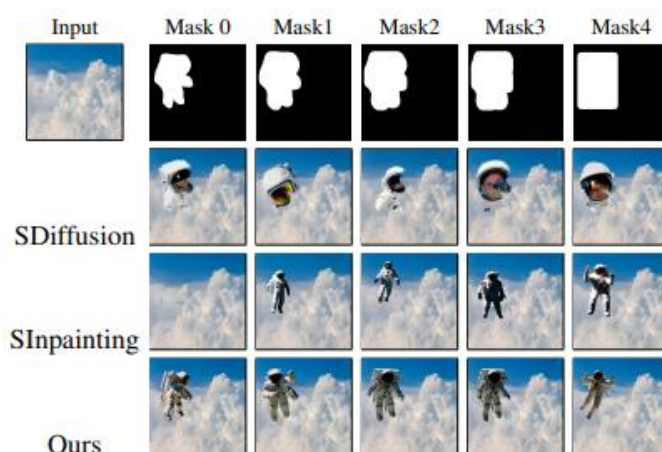


Slika 3.2 korisnike je upitano da odaberu generiranu sliku koja najbolje odgovara maski i unesenom tekstualnom opisu te izgleda najrealističnije [12].

Vizualizacija primjera nadopunjavanja iz istraživanja pokazuje da SmartBrush generira visokokvalitetne objekte i scene, prateći oblik maske i tekst, dok drugi modeli ne uspijevaju pratiti oblik maske. Blended Diffusion i GLIDE ne mogu generirati zadovoljavajuće objekte uz lokalne tekstualne opise. Modeli Stable Diffusion, Stable Inpainting i DALLE-2 pokazuju nešto bolje rezultate, ali s visokom vjerojatnošću nerazumijevanja teksta zbog načina treniranja takvih modela tj. zbog semantičke neusklađenosti između maskiranog, lokalnog područja i teksta koji opisuje cijelu, globalnu sliku (engl. *text misalignment*). Pored nadopunjavanja objekata, SmartBrush podržava i nadopunjavanje scena. Korisnička istraživanja provedena putem Amazon Mechanical Turk pokazuju da više od 50% korisnika preferira rezultate SmartBrusha u svakom od tri

postavljena pitanja, što dodatno potvrđuje prednost predloženog modela u smislu praćenja maske objekta, teksta i ukupne stvarnosti.

U stvarnom svijetu, korisnici neće uvijek pružiti točnu masku objekta kojeg žele nadopuniti. Moguće je da će dostaviti grubu masku, pa SmartBrush omogućuje kontrolu koliko će se nadopunjeni objekt približiti zadanoj maski. Mogu se vidjeti rezultati s različitim vrstama maski (Slika 3.3), koje slijede pravilo zamagljivanja tijekom treninga, odnosno primjene iterativnog Gaussovog zamućivanja kako bi se dobile maske od finih do grubih. Rezultati Stable Diffusiona nisu pod utjecajem vrsta maski jer nije treniran na taj način. S druge strane, rezultati Stable Inpaintinga mijenjaju samo veličinu objekta s veličinom maske, ali ne prate oblik maske. Nasuprot tome, rezultati strogo prate oblik maske kada je dana finija maska, dok otprilike slijede masku ako je dana grublja maska. Za iznimno grubu masku, poput pravokutnog okvira ili pravokutnika (posljednja kolona), dopušta se generiranje bilo gdje unutar njega.



Slika 3.3 Primjeri kontrole preciznosti s promptom „astronaut“ [12].

U zaključku istraživanja predstavljenog u radu, razvijen je inovativan pristup rješavanju izazova s kojima se suočavaju postojeći modeli za nadopunjavanje slika uz vođenje tekstem i oblikom. Kroz korištenje smjernica teksta i oblika iz skupa podataka segmentacije, istraživači su uspješno objasnili problem nepodudaranja teksta. Dodatno, predložena je precizna kontrola generacije putem stvaranja različitih razina maski, omogućujući korisnicima da određuju koliko će precizno generirani objekt pratiti zadanu masku. Rezultati kvantitativnih i kvalitativnih evaluacija potvrđuju nadmoćnost predložene metode u odnosu na postojeće pristupe. Važno ograničenje predstavlja situacija s velikom sjenom, gdje sjena objekta prelazi masku objekta, npr., sjena osobe može biti vrlo dugačka dok okvir obično ne uspije pokriti cijelu sjenu. Tim istraživanjem otvaraju se smjerovi za daljnji razvoj i poboljšanje.

3.2. Scene Composer

Postojećim modelima generiranja slika nedostaje fleksibilnost u podržavanju punog kreativnog tijeka rada. Scene Composer ima za cilj generirati slike iz semantičkih rasporeda s različitim razinama preciznosti, omogućujući veću kontrolu u kreativnom procesu [13]. Semantički raspored predstavljen je kao skup regija s tekstualnim opisima slobodnog oblika, a razina preciznosti kontrolira koliko dobro generirani objekt treba odgovarati navedenom obliku. Kodiranje rasporeda otvorene domene predstavlja inovativni pristup u kojem se predstavlja mapa značajki teksta kako bi se kodirali semantički rasporedi. U istraživanju iz ovog članka su opisani problemi s izazovima rukovanja ručno nacrtanim rasporedima, te se uvodi precizno kodirana piramida maske za modeliranje preciznosti rasporeda. Nedostatak skupova podataka velikih razmjera rješava se prikupljanjem podataka iz dva izvora, parova slika – tekst velikih razmjera i manjeg skupa podataka pseudo rasporeda/slike, koristeći otkrivanje i segmentaciju objekata temeljenih na tekstu. Arhitektura modela uključuje piramidu precizno kodiranih maski koja predstavlja svaku masku u rezolucijama do 2^{c_i} , gdje je c_i razina preciznosti. Najmanja razina, za $c_i=0$, predstavlja pretvaranje teksta u sliku (tj. pridružena maska se zanemaruje), najveća razina precizno pretvaranje segmentirane maske u sliku, dok međurazine opisuju koliko točno treba generirani sadržaj pratiti pripadajuću masku. Također, tu je piramida tekstualnih značajki koja pruža fleksibilan kontrolni mehanizam predstavljanja tekstualnih značajki u različitim rezolucijama. Model vođene difuzije na više skala koristi se za generiranje slika iz piramide značajki teksta. Doprinosi ovog istraživanja obuhvaćaju unificirani okvir za sintezu slike temeljenu na difuziji iz semantičkih rasporeda s kontrolom preciznosti. Istraživanje također donosi novi skup podataka o rasporedu koji crta korisnik u stvarnom svijetu, a eksperimenti detaljno pokazuju učinkovitost predloženog modela. Povezani radovi raspravljaju o dubokim generativnim modelima, tehnikama generiranja slike i radovima na sintezi uvjetno generirane slike, naglašavajući razlike u odnosu na postojeće pristupe. Eksperimentalni rezultati obuhvaćaju procjenu predložene metode za generiranje rasporeda u sliku otvorene domene, generiranje teksta u sliku te generiranje segmentacije u sliku. Prijavljeni su kvantitativni rezultati, uključujući rezultate prostorne sličnosti, CLIP rezultate i FID. Sveukupno, istraživanje je usredotočeno na pružanje fleksibilnog modela za sintezu slike koji se može kontrolirati i prilagoditi različitim razinama preciznosti u semantičkim rasporedima, nudeći podršku za različite faze kreativnog procesa. Na Slici 3.4 mogu se vidjeti rezultati s različitim razinama preciznosti.

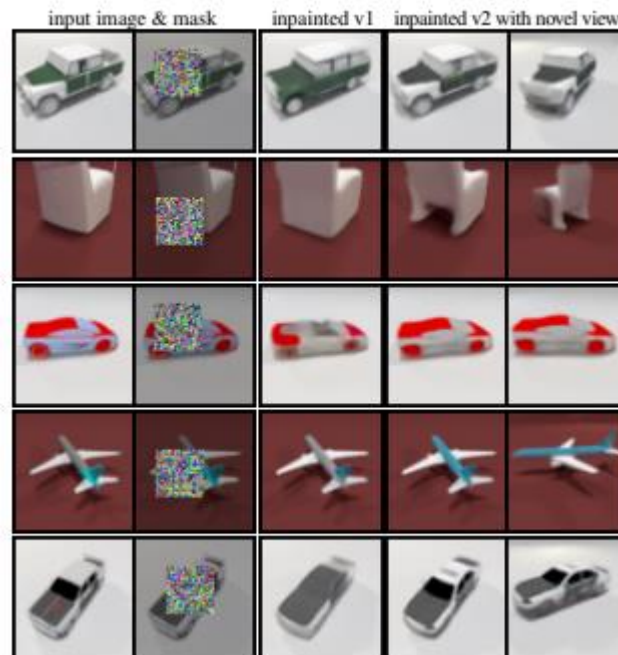


Slika 3.4 Rezultati s različitim razinama preciznosti. Za svaki ulazni raspored, uzorkovane su slike počevši od istog šuma, tako da slike na različitim razinama preciznosti mogu imati slične stilove [13].

3.3. RenderDiffusion

Do sada, modeli difuzije slika nisu podržavali zadatke potrebne za razumijevanje 3D-a, poput generiranja 3D-a koji je u skladu s pogledom ili rekonstrukcije objekata iz jednog pogleda. RenderDiffusion, prvi je difuzijski model za generiranje i zaključivanje 3D-a, treniran korištenjem samo monokularnih 2D slika [14]. RenderDiffusion se može koristiti za razumijevanje 3D scene iz slike, za 3D uređivanje pomoću 2D retuširanja i za generiranje 3D scena. Pokazane su konkurentne performanse na zadacima uzorkovanja i zaključivanja, kako u smislu kvalitete, tako i raznolikosti rezultata. Ova metoda trenutno ima nekoliko ograničenja. Prvo, generirane slike još uvijek zaostaju za GAN-ovima u nekim slučajevima, vjerojatno zbog izlaza koji su ponekad manje oštri; međutim, ovaj model prvi je 3D-difuzijski model treniran s 2D slikama, i postoje inženjerske tehnike koje bi mogle poboljšati kvalitetu, poput modela za povećanje rezolucije. Drugo, uvedena je regularizacija korištenjem uzorkovanja bodovne funkcije za funkciju gubitka (engl. *score distillation sampling loss*, SDS loss) kako bi se spriječilo učenje trivijalne geometrije na bazi sa stvarnim ljudskim licima (Flickr-Faces-HQ, FFHQ) i bazi s mačjim licima (Animal FacesHQ, AFHQ), što rezultira gubitkom vjernosti u generiranim 3D modelima. Očekuje se da će ovo manje biti problem kako se SDS metode poboljšavaju. Treće, zahtijevalo se da slike za treniranje uključuju ekstrinzične parametre kamere, a reprezentacija 3D scene s tri 2D ravnine značajki (engl. *triplane*) postavljeni su u globalni koordinatni sustav, što ograničava generalizaciju između raspoređivanja objekata. Ovo ograničenje može se riješiti korištenjem dostupnih metoda procjene položaja ili renderiranjem svega u pogledu kamere. Konačno, prema istraživanju, željelo bi se podržati uređivanje objekata i uređivanje materijala, omogućujući tako izražajni 3D-svjesni radni

tok uređivanja 2D slika. Na slici ispod prikazano je nadopunjavanje unutar slike svjesno 3D prostora.



Slika 3.5 Nadopunjavanje unutar slike svjesno 3D prostora [14].

3.4. Imagen Editor

Imagen Editor koristi se za lokalizirano nadopunjavanje slike iz tekstualnog opisa [15]. Primjere uređivanja pomoću programa Imagen Editor možemo vidjeti na Slici 3.6. Tekstom vođeno uređivanje slika može imati transformirajući utjecaj u podršci kreativnim aplikacijama. Ključni izazov je generiranje uređivanja koja su vjerna tekstualnim uputama i istovremeno konzistentna s ulaznim slikama. Imagen Editor, kaskadni model difuzije, izgrađen je poboljšavanjem modela Imagen na tekstualno vođenoj slikovnoj slikovnom nadopunjavanju. Uređivanja Imagen Editora vjerna su tekstualnim uputama, postignuta korištenjem detektora objekata za predlaganje maski popune tijekom treninga. Za poboljšanje kvalitativne i kvantitativne evaluacije uvodi se EditBench, sustavna provjera za tekstualno vođenu slikovnu nadopunu. EditBench ocjenjuje popune slikovne nadopune na prirodnim i generiranim slikama koje istražuju objekte, atribute i scene. Kroz opsežnu ljudsku evaluaciju na EditBench-u, utvrđeno je da maskiranje objekta tijekom treninga dovodi do univerzalnih poboljšanja u poravnanju teksta i slika, tako da se Imagen Editor preferira u odnosu na DALL-E 2 i Stable Diffusion. Također, zaključeno je da su ovi modeli bolji u prikazu objekata nego teksta, te bolje koriste atributima materijala/boje/veličine nego atributima broja/oblika.



Slika 3.6 Niz uređivanja pomoću programa Imagen Editor [15].

3.5. Inst-Inpaint

Nadopunjavanje slike (engl. *image inpainting*) odnosi se na zadatak uklanjanja neželjenih objekata i/ili popunjavanja nedostajućih područja unutar slike. Smatra se izuzetno izazovnim jer bi nadopunjena područja na rezultirajućim slikama trebala biti u harmoniji s ostatkom slike. Pristupi nadopunjavanju trebaju uzeti u obzir globalni kontekst slike prilikom dodavanja nedostajućih piksela. Tradicionalno, postojeća istraživanja koriste binarne maske kako bi istakle nedostajuća područja. Kao i u mnogim zadacima poboljšanja slika, nedavni pristupi nadopunjavanju slika koriste modele temeljene na dubokom učenju opremljene generativnim priorima i formuliraju zadatke nadopunjavanja kao problem nadziranog učenja. Drugim riječima, trenirajući podaci uključuju parove ulaznih i izbrisanih slika koje djeluju kao izvorni i ciljni podaci, respektivno, a modeli se općenito treniraju koristeći cilj temeljen na rekonstrukciji. Iako je uočeno kontinuirano poboljšanje performansi pristupa nadopunjavanju tijekom proteklih godina, ovi pristupi i dalje zahtijevaju maske koje korisnici specificiraju prilikom testiranja modela. Ove maske mogu biti nacrtane u obliku tankih ili debelih poteza kistom ili crtanjem okvira, ali crtanje ovih maski često je zamorno i podložno pogreškama.

Nedavno su generiranje i uređivanje slika na temelju teksta privukli puno pažnje. Pogotovo kada su trenirani na velikim skupovima podataka slika i teksta, modeli poput DALL·E2, Stable Diffusion, Imagen i Parti pokazuju izuzetne generativne sposobnosti dok učinkovito bilježe kompozicijsku strukturu slika. Osim toga, korištenje teksta omogućuje prirodniju i korisnički jednostavniju kontrolu nad slikama. Štoviše, istraživanja su pokazala da se ovi pristupi lako mogu prilagoditi nadopunjavanju slika dodatnim uvjetovanjem modela binarnim maskama i reći što treba ispuniti unutar maske. Međutim, taj pristup bolje odgovara zamjeni atributima objekta ili dodavanju novog elementa scene, ali možda neće tako dobro funkcionirati za uklanjanje objekata. Zbog toga se u radu [16] predlaže nadopunjavanje slika instrukcijama, u kojem se neželjeni objekti specificiraju samo putem tekstualnih uputa, bez potrebe za binarnim maskama. Za rješavanje tog problema koristi se skup podataka GQA, te se modelira nova baza za generiranje podataka kako

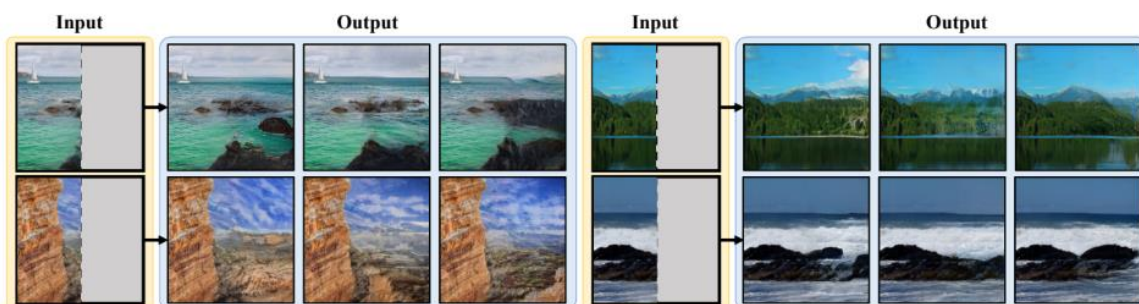
bi se stvorio novi skup podataka za procjenu, koji se naziva GQA-Inpaint. Drugo, dizajnira se jednostavna duboka mreža za nadopunjavanje, Inst-Inpaint, koja može uklanjati objekte sa slika koristeći upute dane kao tekst. Za razliku od postojećih metoda nadopunjavanja, ovaj model ne zahtijeva maske, niti eksplicitno predviđanje maski. Na Slici 3.7 prikazano je nadopunjavanje slika instrukcijama na nekoliko različitih primjera.



Slika 3.7 Nadopunjavanje slika instrukcijama [16].

4. Nadopunjavanje izvan slike (engl. outpainting)

Nadopunjavanje izvan slike (engl. *outpainting*) zahtijeva semantički dosljedno proširenje ulazne slike izvan njezinog dostupnog sadržaja [17]. Za razliku od nadopunjavanja unutar slike, koji se bavi popunjavanjem nedostajućih piksela na koherentan način sa susjednim pikselima, nadopunjavanje unutar slike može biti postignuto na različite načine, budući da je problem manje ograničen susjedstvom piksela. Na primjer, ako je prikazana slika planina, moguće je mentalno ekstrapolirati i zamisliti okolne šumske ili snježne scene, pretpostaviti jezero ispod padine brda i vizualizirati stijene blizu oceana. Postojeće metode nadopunjavanja unutar slike postavljaju problem kao uvjetni zadatak prijevoda slike u sliku, često generirajući repetitivne strukture i teksture repliciranjem dostupnog sadržaja na ulaznoj slici. Prema istraživanju, problem se formulira iz perspektive inverzije generativnih suparničkih mreža (GAN). Generator renderira mikroregije uvjetovane njihovim zajedničkim latentnim kodom, kao i njihovim pojedinačnim pozicijama na slici. Za nadopunjavanje slike traže se više latentnih kodova koji ne samo da obnavljaju dostupne regije, već i stvaraju raznoliko nadopunjavanje putem generacije bazirane na regijama. To dovodi do bogatije strukture i sadržaja u proširenim regijama. Osim toga, prema istraživanju, formulacija omogućuje nadopunjavanje izvan slike uvjetovano na kategorički unos, pružajući prilagodljivu kontrolu korisnicima. Pokazano je da predložena metoda ima povoljne rezultate u usporedbi s postojećim metodama nadopunjavanja unutar i izvan slike, s većom vizualnom kvalitetom i raznolikošću. Na Slici 4.1 prikazano je nadopunjavanje izvan slike na više različitih primjera.

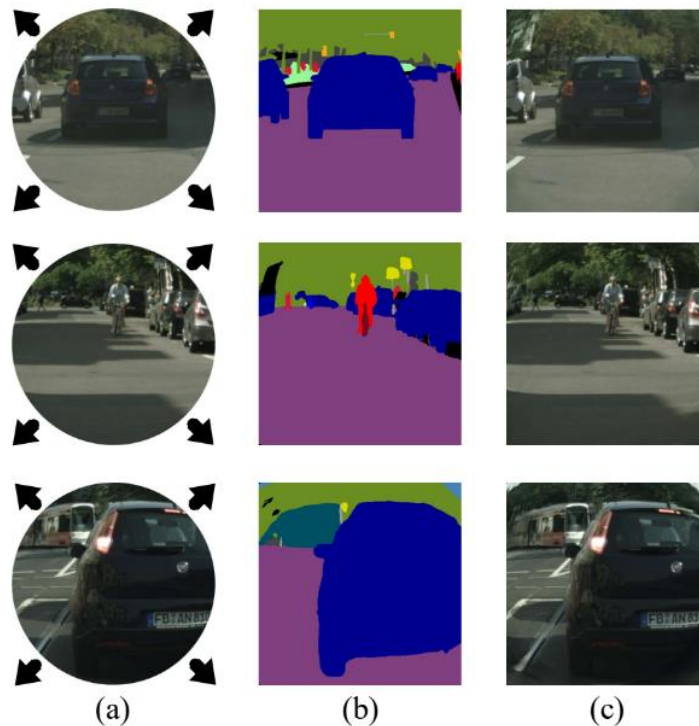


Slika 4.1 Nadopunjavanje izvan slike na više različitih primjera, predložena metoda može tražiti različita rješenja za određeni ulaz, postižući visoku raznolikost rezultata bez žrtvovanja kvalitete generiranja [17].

4.1. FishDreamer

FishDreamer je napredna metoda razvijena za semantičku nadopunu širokokutnih kamera "riblje oko" (engl. *fisheye*), koristeći tehnike nadopunjavanja izvan slike i unifikacije slika [17]. Cilj FishDreamera je poboljšati razumijevanje i rekonstrukciju slika snimljenih kamerama "riblje

oko" koje često imaju pogreške od distorzija i ograničenih vidnih polja. Ova metoda kombinira semantičku nadopunu i nadopunu slike kako bi generirala slike koje nadilaze ograničenja takvog formata, stvarajući cjelovite i prirodne slike. Istraživački rad pod imenom "Fisheye Semantic Completion" (FSC) [17], opisuje gustoću tekstone, strukturu i semantiku širokokutne slike koje se zaključuju čak i izvan vidnog polja senzora (engl. *field of view*, FoV). Fisheye kamere imaju veći FoV od običnih kamera s rupicom, ali njihov jedinstveni model snimanja prirodno dovodi do slijepe zone na rubu ravnine slike. Ovo je suboptimalno za sigurnosno kritične primjene jer važni zadaci percepcije, poput semantičke segmentacije, postaju vrlo izazovni unutar slijepe zone. Prethodni radovi razmatrali su nadopunjavanje izvan FoV i segmentaciju unutar FoV zasebno. Međutim, primijećeno je da su ova dva zadatka zapravo blisko povezana. Kako bi se procijenila čvrsto isprepletena potpuna slika "riblje oko" i scenska semantika, predstavljen je novi FishDreamer koji se oslanja na uspješne vizualne transformer mreže (engl. *visual transformer*, ViT) poboljšane novim polarno-svjesnim modulom križne pažnje (engl. *polar-aware cross attention module*, PCA) kako bi iskoristile gusto kontekstualno okruženje i vodile generaciju semantički dosljednog sadržaja uzimajući u obzir različite polarne distribucije. Istraživanje je pokazalo da predloženi FishDreamer nadmašuje metode koje rješavaju svaki zadatak zasebno i nadmašuje alternativne pristupe u okviru rada. FishDreamer koristi PCA modul koji poboljšava interpretaciju i rekonstrukciju scena. Model je testiran na nekoliko skupova podataka, uključujući Cityscapes i KITTI360, gdje je pokazao superiorne performanse u odnosu na postojeće metode poput osnovnog modela bez PCA i modela. Jedna od ključnih komponenti FishDreamera je glava za semantičko nadopunjavanje (engl. *semantic completion head*), koji se fokusira na semantičku interpretaciju slike, te glava za nadopunjavanje slike (engl. *image completion head*), koji se bavi prostornom rekonstrukcijom slike izvan granica formata širokokutne slike. Ove komponente zajedno omogućuju FishDreameru da nadmaši postojeće metode u stvaranju vizualno uvjerljivih i semantički točnih slika. Prikaz Ilustracije Fisheye Semantic Completiona nalazi se na Slici 4.2 u nastavku.



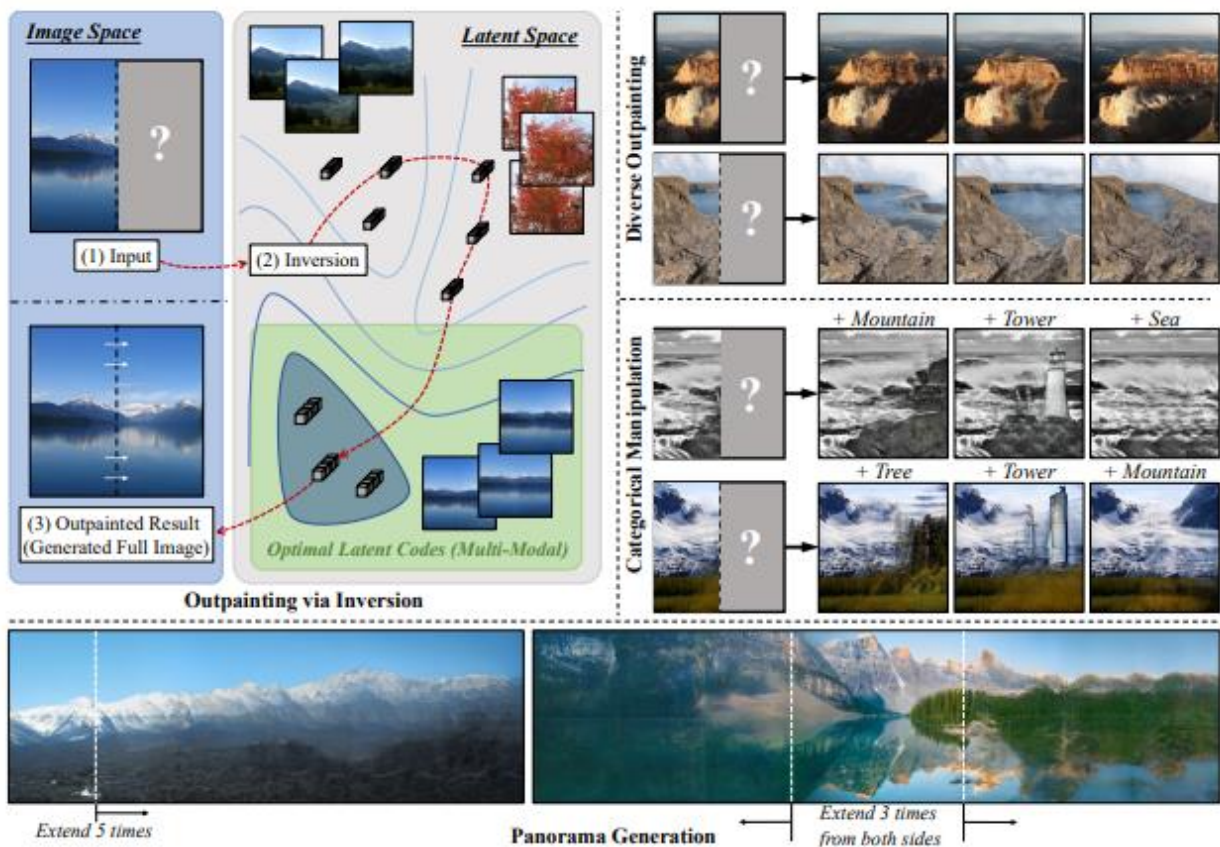
Slika 4.2 Ilustracija Fisheye Semantic Completiona [17].

4.2. InOut

Postojeće metode nadopunjavanja slike često generiraju ponavljajuće strukture i teksture ponavljanjem sadržaja dostupnog u ulaznoj slici. U radu iz kojeg je uzeta referenca formuliran je problem iz perspektive inverznih generativnih suparnički mreža (inverzni GAN) [19]. Generator prikazuje mikroregije uvjetovane njihovim zajedničkim latentnim kodom kao i njihovim pojedinačnim položajima u slici. Nadopunjavanje slike postiže se traženjem više latentnih kodova koji ne samo da obnavljaju dostupne regije već i sintetiziraju raznolika nadopunjavanja generiranjem mikroregija. To dovodi do bogatijih struktura i sadržaja u nadopunjenim područjima. Nadalje, formulacija omogućava nadopunjavanje uvjetovano kategorijskim unosom, čime se omogućava fleksibilna korisnička kontrola. Opsežni eksperimentalni rezultati pokazuju da predložena metoda nadmašuje postojeće metode nadopunjavanja slike, pružajući višu vizualnu kvalitetu i raznolikost. Nadopunjavanje slike omogućuje različite aplikacije za stvaranje sadržaja, kao što su uređivanje slika korištenjem ekstrapoliranih regija, generiranje panoramskih slika i prošireno iskustvo u virtualnoj stvarnosti. Nedavni napredci u popunjavanju slike ne rješavaju izravno problem nadopunjavanja, jer popunjavanje ima više konteksta za rad - nedostajući pikseli imaju veći broj dostupnih okolnih piksela koji služe kao granični uvjeti i pružaju ključne smjernice za popunjavanje. Nasuprot tome, problem nadopunjavanja izvan slike može se osloniti samo na kontekst dostupne slike, s malim brojem piksela blizu granice dostupnih kao granični uvjet. Slična

je analogija između interpolacije videa i predikcije videa, gdje interpolacija radi s postojećim događajima, dok predikcija pokušava modelirati više budućnosti. U spomenutom istraživačkom radu rješava se problem nadopunjavanja slike inverznim generativnim suparničkim mrežama (inverzni GAN). Prvo se proširuje generator temeljen na StyleGAN2 za generiranje u koordinatno uvjetovanom načinu i neovisno generirani prostorno dosljedne mikroregije. Svaka mikroregija dijeli globalni latentni kod s ostatkom mikroregija u slici, dok ima jedinstvenu oznaku koordinata. Nadopunjavanje se formulira kao pronalaženje optimalnih latentnih kodova za dostupne ulazne mikroregije, nakon čega slijedi generiranje željenih regija pružanjem odgovarajuće koordinatne uvjetovanosti. Traži se latentni kod pomoću inverzne GAN mreže koja pronalazi više latentnih kodova proizvodeći raznolike nadopunjene regije, otključavajući raznolikost u izlazu.

Evaluacija predložene metode na Place365 i Flickr-Scenery bazama pokazuje da metoda nadmašuje postojeće metode u smislu vizualne kvalitete i raznolikosti. Koristeći metodu Fréchet Inception Distance (FID) i Inception Score (IS), demonstrirano je da distribucije generiranih slika tom metodom bolje odgovaraju stvarnim distribucijama nego distribucije generirane postojećim metodama. Korištenjem korisničkih studija i kvantitativnih eksperimenata, pokazuje se da korisnici preferiraju rezultate ove metode zbog bolje sinteze složenih struktura i novih objekata. Prikaz funkcije raznolikog nadopunjavanja slike obzirom na predloženi algoritam nalazi se na Slici 4.3.



Slika 4.3 (gore-lijeva) - obzirom na ulaznu sliku i istrenirani generator, predloženi algoritam traži

latentne kodove koji mogu generirati slike koje sadrže ulaznu sliku. Na taj način možemo prirodno postići (gore-desno) raznoliko nadopunjavanje slike, (sredina-desno) kategorijsku manipulaciju za područje nadopunjavanja, i (dolje) generirati panoramu s bogatom i složenom strukturom [19].

5. Subjektivna evaluacija kvalitete slike

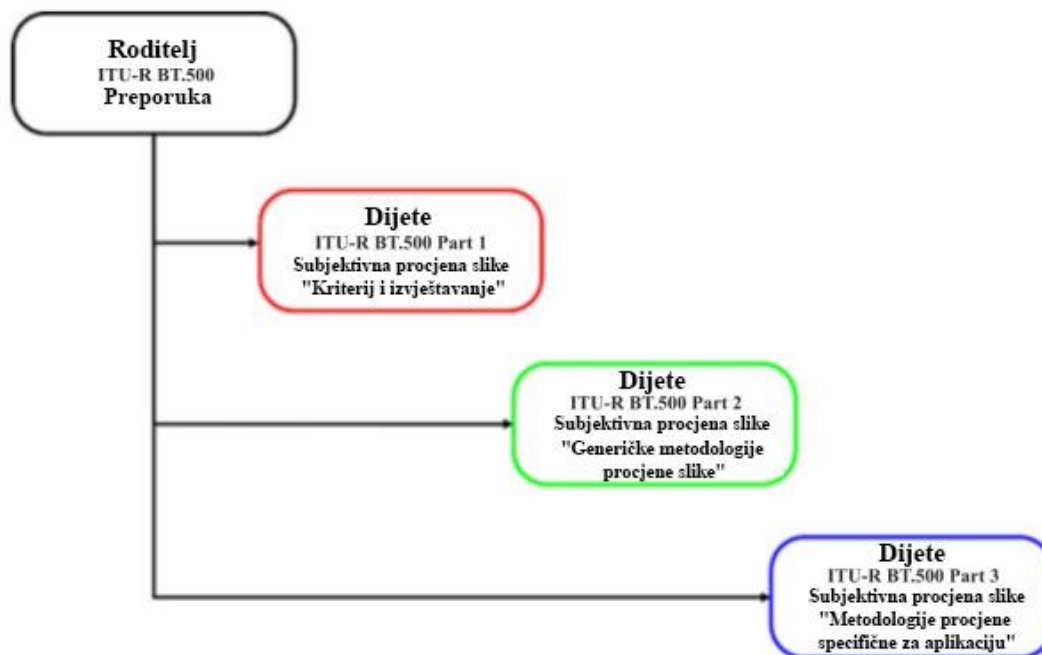
Subjektivna evaluacija kvalitete slike je metoda temeljena na ljudskom opažanju i mišljenju. Budući da su ljudski promatrači krajnji korisnici većine multimedijских aplikacija, najtočniji i najpouzdaniji način procjene kvalitete slika je putem subjektivne evaluacije. Međutim, subjektivne evaluacije su skupe i zahtijevaju puno vremena, što ih čini nepraktičnim za stvarne primjene. Nadalje, subjektivni eksperimenti su dodatno komplicirani mnogim čimbenicima, uključujući udaljenost gledanja, uređaj za prikaz, uvjete osvjetljenja, vidnu sposobnost promatrača i njihovo raspoloženje. Stoga je potrebno dizajnirati matematičke modele koji su sposobni predvidjeti procjenu kvalitete prosječnog ljudskog promatrača. U subjektivnom testiranju, grupa ljudi daje svoje mišljenje o kvaliteti svake slike. Kako bi se dobili pouzdani rezultati, koriste se međunarodni standardi poput ITU-R BT.500-15 [18], koji daje smjernice za uvjete gledanja, način provođenja subjektivnih eksperimenata, materijale za testiranje i prezentaciju rezultata. Ove smjernice osiguravaju dosljednost i ponovljivost subjektivnih procjena. U nastavku će biti opisane neke od standardiziranih subjektivnih metoda procjene kvalitete slike (IQA). Jednopodražajno kategorizirano ocjenjivanje metoda je u kojoj se testne slike prikazuju na ekranu određeno vrijeme, nakon čega nestaju s ekrana i promatrači ocjenjuju njihovu kvalitetu na apstraktnoj skali koja sadrži jednu od pet kategorija: izvrsno, dobro, zadovoljavajuće, loše ili vrlo loše. Sve testne slike prikazuju se nasumično. Kako bi se izbjegli kvantizacijski artefakti, neke metode koriste kontinuirane umjesto kategoriziranih skala. Dvopodražajno kategorizirano ocjenjivanje metoda je slična prethodnoj, ali se u ovom slučaju prikazuju i testne i referentne slike određeno vrijeme. Nakon toga, slike nestaju s ekrana i promatrači ocjenjuju kvalitetu testne slike prema opisanoj apstraktnoj skali. Kod subjektivnog ocjenjivanja parnim uspoređivanjem prisilnim izborom dvije slike istog prizora prikazuju se promatračima. Nakon toga, traži se da odaberu sliku više kvalitete. Promatrači uvijek moraju odabrati jednu sliku, čak i ako obje slike nemaju razlike. Ova metoda zahtijeva više pokušaja kako bi se usporedili svi parovi uvjeta. Ocjenom parne sličnosti promatrači su zamoljeni ne samo da odaberu sliku više kvalitete, već i da označe razinu razlike između njih na kontinuiranoj skali. Izravno korištenje rezultata ocjenjivanja kao što su izvrsno, dobro, zadovoljavajuće itd. za ocjene kvalitete može biti nepouzdana jer promatrači mogu dodijeliti različite ljestvice kvalitete za svaki prizor i tip izobličenja. Z-ocjena (engl. *Z-score*) koristi se kako bi se lakše usporedila mišljenja svakog promatrača o kvaliteti slika, tj. koristi se linearna transformacija koja čini srednju vrijednost i varijancu jednakima za sve promatrače. Rezultat takve transformacije naziva se Z-ocjena. Razlika prosječne ocjene (engl. *difference mean opinion score*, DMOS) umjesto izravne primjene rezultata ocjenjivanja (engl. *mean opinion score*, MOS), moderne IQA mjere kvalitete slike (engl. *image quality assessment*, IQA) koriste razlike u kvaliteti

između slika. DMOS se definira kao razlika između izravne ocjene kvalitete referentnih i testnih slika. Subjektivne metode procjene kvalitete pružaju točna i pouzdana mjerenja kvalitete vizualnih signala. Međutim, ove metode imaju različite nedostatke koji ograničavaju njihovu primjenu. Zahtijevaju mnogo vremena i skupe su zbog provođenja eksperimenata s mnogim promatračima. Ne mogu se integrirati u aplikacije u stvarnom vremenu kao što su sustavi za kompresiju i prijenos slika. Njihovi rezultati uvelike ovise o fizičkom stanju i emocionalnom raspoloženju promatrača, kao i o drugim čimbenicima kao što su uređaj za prikaz i uvjeti osvjetljenja. Stoga je potrebno dizajnirati matematičke modele koji su sposobni predvidjeti perceptivnu kvalitetu vizualnih signala na dosljedan način sa subjektivnim evaluacijama.

5.1. Preporuka ITU-R BT.500-15

Preporuka ITU-R BT.500-15 je dokument Međunarodne telekomunikacijske unije (ITU) koji pruža sveobuhvatne metodologije za subjektivnu procjenu kvalitete televizijskih slika [19]. Ova preporuka, koja se redovito ažurira, posljednji put u svibnju 2023. godine, igra ključnu ulogu u standardizaciji procjene kvalitete slike u radiokomunikacijama, posebno u kontekstu televizijskih usluga. ITU-R BT.500-15 sastoji se od tri glavna dijela. Prvi dio opisuje sveukupne zahtjeve za provođenje subjektivne procjene kvalitete slike, uključujući opće metode testiranja, ljestvice ocjenjivanja i preporučene uvjete gledanja. Drugi dio pruža detaljan pregled preporučenih metodologija koje se mogu koristiti prilikom izvođenja subjektivnih procjena, dok treći dio opisuje metodologije specifične za različite formate slika i aplikacije, temeljem specifikacija iz prva dva dijela. Obzirom na ranije spomenute izazove kod subjektivnog ocjenjivanja kvalitete slike, postoji potreba za razvojem matematičkih modela koji mogu predvidjeti procjenu kvalitete slike prosječnog ljudskog promatrača na dosljedan način. U preporuci ITU-R BT.500-15 opisane su različite standardizirane metode subjektivne procjene kvalitete slike. Jedna od ključnih metoda je jednopodražajno kategorizirano ocjenjivanje, gdje se testne slike prikazuju na ekranu određeno vrijeme, nakon čega nestaju, a promatrači ocjenjuju njihovu kvalitetu na apstraktnoj ljestvici s pet kategorija: izvrsno, dobro, zadovoljavajuće, loše ili vrlo loše. Sve testne slike se prikazuju nasumično kako bi se izbjegla pristranost. Druga važna metoda je dvopodražajno kategorizirano ocjenjivanje, gdje se prikazuju i testne i referentne slike određeno vrijeme, a zatim promatrači ocjenjuju kvalitetu testne slike prema apstraktnoj ljestvici. Dodatno, metoda usporedbe po parovima s prisilnim odabirom uključuje prikazivanje dvije slike iste scene promatračima, koji moraju odabrati sliku s višom kvalitetom, čak i ako nema razlike između njih. U usporedbi po parovima sa ocjenom sličnosti, promatrači ne samo da odabiru sliku s višom kvalitetom, već i naznačuju razinu razlike između njih na kontinuiranoj ljestvici. Modernije metode uključuju razliku prosječne ocjene mišljenja (DMOS), gdje se umjesto izravne primjene rezultata

ocjenjivanja koristi razlika u kvaliteti između referentne i testne slike. Konačno, Z-ocjena koristi linearne transformacije kako bi srednja vrijednost i varijanca bile jednake za sve promatrače, omogućujući lakšu usporedbu mišljenja. Subjektivne metode procjene kvalitete pružaju točna i pouzdana mjerenja kvalitete vizualnih signala, ali njihova primjena je ograničena visokim troškovima i vremenskim zahtjevima. Preporuka ITU-R BT.500-15 pruža temeljne smjernice za takve procese, osiguravajući dosljednost i točnost procjena kvalitete slike širom svijeta. Prikaz strukture preporuke ITU-R BT.500 na Slici 5.1.

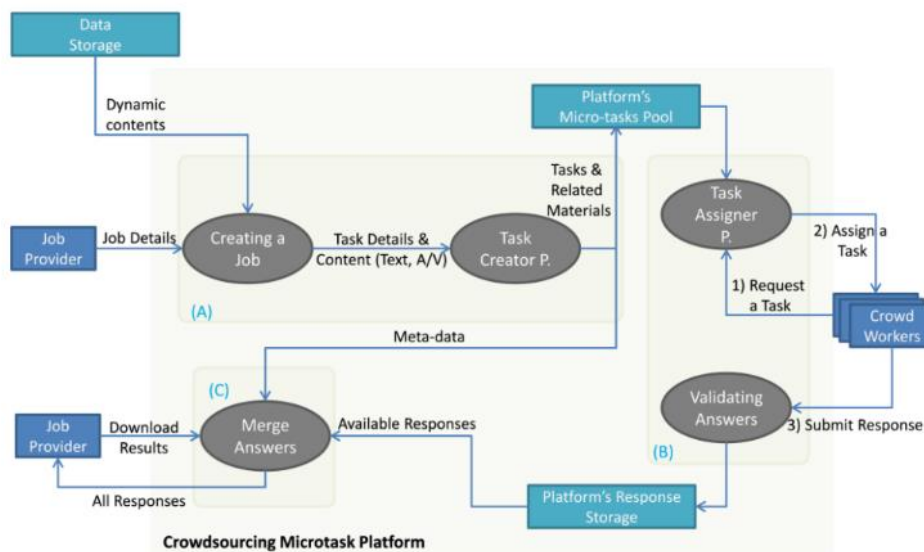


Slika 5.1 Preporuka ITU-R BT.500 struktura [19].

5.2. Masovna ispitivanja (eng. *crowdsourcing*)

Masovna ispitivanja, pojam uveden 2006. godine, spaja masu (engl. *crowd*) i korištenje vanjskih usluga (engl. *outsourcing*), odnosi se na praksu izdvajanja poslova koji su tradicionalno obavljani određenim agentom (obično zaposlenikom) nedefiniranoj, obično velikoj grupi ljudi putem otvorenog poziva. Ova metoda koristi potencijalnu sposobnost velike grupe ljudi povezane putem interneta za obavljanje određenog zadatka. Tijekom vremena, razvile su se različite online zajednice pod okriljem masovnih ispitivanja kako bi riješile različite probleme primjenjujući različite metode suradnje. Primjeri u različitim područjima uključuju Wikipediju, Linux, Stack Overflow i Amazon Mechanical Turk. Sustavi za masovna ispitivanja mogu se klasificirati prema tome tko čini tu zajednicu, što rade, zašto i kako to rade. Važan aspekt je jesu li članovi mase neplaćeni volonteri ili plaćeni za svoje usluge. Komercijalne platforme s plaćenim članovima šire

se u različitim dimenzijama. Tipično djeluju kao posrednici i pružaju zadatke koje poslodavci daju masi, bilo u obliku natjecanja ili otvorenog poziva. Poseban slučaj je masovno ispitivanje mikro-zadataka, koji pruža iznimnu priliku nudeći visoko skalabilnu, na zahtjev i jeftinu radnu snagu za obavljanje složenih zadataka koji se mogu podijeliti na niz kratkih i jednostavnih online zadataka poput anotacija i prikupljanja podataka. Uobičajeno je da radnici u masi budu plaćeni na osnovi obavljenog rada, a zadaci se nude putem platformi za masovna ispitivanja, gdje poslodavci stvaraju poslove koji su zbirka zadataka koje mogu obaviti jedan ili više članova mase u zamjenu za novčane nagrade. Masovna ispitivanja kvalitete multimedijjskih sadržaja, iako nisu direktna implementacija laboratorijskih metoda, zahtijevaju prilagodbu za internetsko okruženje. Pri prijenosu pasivnog laboratorijskog eksperimenta na masovna ispitivanja, važno je uzeti u obzir razlike u opremi radnika i pripremiti strukturu baze podataka [20]. Tehnički izvještaj ITU-T PSTR-CROWDS iz 2018. godine opisuje osnovne koncepte masovnih ispitivanja, njegovu primjenu u procjeni kvalitete te opće principe subjektivne metodologije procjene kvalitete medija pomoću mikro-zadataka masovnih ispitivanja. Mikro-zadaci nude brze, jeftine i skalabilne pristupe izdvajanjem zadataka velikom broju sudionika. Masovno ispitivanje omogućava pristup raznolikoj populaciji sudionika u stvarnim okruženjima, čime se smanjuje trošak i vrijeme potrebno za provođenje eksperimenata [21]. Eksperiment treba biti ograničen veličinom sesije testa za masovno ispitivanje, a valjanost i pouzdanost prikupljenih podataka kontroliraju se kroz različite metode. Priprema za test uključuje jasne upute, obuku radnika i ocjenjivanje valjanosti, dok se tijekom izvođenja testa preporučuju transparentnost, pravedna naknada i pilot testiranje. Za analizu se preporučuju standardne statističke metode i evaluacija pouzdanosti odgovora radnika, uz uvođenje provjera valjanosti tijekom pripreme, izvođenja i pregleda podataka [20]. Prikaz općeg tijek rada platformi za mikro zadatke masovnog ispitivanja na Slici 5.2.



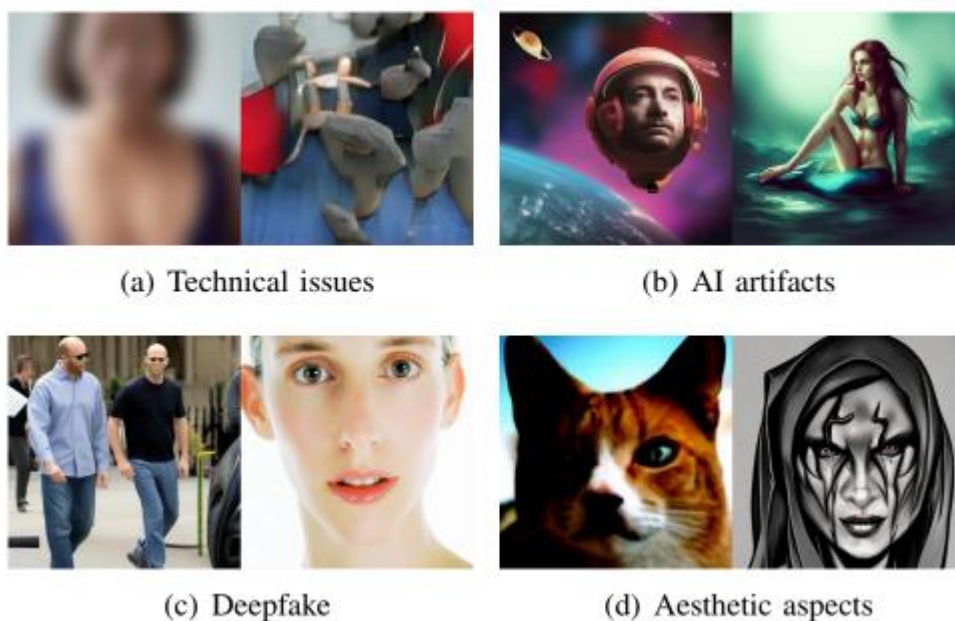
Slika 5.2 Opći tijek rada platformi za mikro zadatke masovnog ispitivanja [20].

5.3. AGIQA-3K baza podataka

AGIQA-3K je sveobuhvatna baza podataka namijenjena procjeni kvalitete slika generiranih umjetnom inteligencijom (AI), s posebnim naglaskom na generativne modele tekst-u-sliku [22]. Ova baza podataka je osmišljena kako bi pokrila široki spektar kvalitete slika proizvedenih različitim AI modelima, uključujući generativne suparničke mreže (GAN), autoregresivne modele i difuzijske modele. S ukupno 2 982 slike, generirane pomoću šest različitih modela, AGIQA-3K pruža detaljne subjektivne procjene percepcije i usklađenosti teksta i slike. Kako bi se osigurala točnost procjene kvalitete slika, AGIQA-3K koristi subjektivne eksperimente provedene u standardiziranom laboratorijskom okruženju. Sudionici ovih eksperimenata ocjenjuju slike na temelju nekoliko kriterija, uključujući tehničke probleme, artefakte koje stvara AI, estetske aspekte i autentičnost slike. Ove subjektivne ocjene omogućuju detaljnu analizu performansi različitih generativnih modela. Jedan od glavnih izazova u procjeni kvalitete slika generiranih AI modelima je velika varijabilnost u kvaliteti slika, koja ovisi o modelu i postavkama korištenim tijekom generiranja. Prikaz nekih uobičajenih izobličenja generiranih umjetnom inteligencijom uočava se na Slici 5.4. AGIQA-3K nastoji riješiti ovaj problem pružajući detaljne subjektivne ocjene koje odražavaju percepciju ljudskog vizualnog sustava (engl. *human visual system*, HVS). To omogućuje razvoj i evaluaciju modela za procjenu kvalitete slika koji su usklađeni s ljudskom percepcijom. Baza podataka AGIQA-3K koristi mehanizam prikupljanja tekstualnih opisa koji kombinira stvarne tekstualne opise i one dizajnirane od strane ljudi, osiguravajući širok raspon ulaznih sadržaja u T2I generacijskim zadacima. Eksperimenti se provode prema standardu ITU-R BT.500-13, što osigurava konzistentnost i valjanost prikupljenih subjektivnih ocjena. AGIQA-3K pruža vrijedne resurse za istraživače i programere koji žele poboljšati kvalitetu generativnih AI modela. Omogućuje im optimizaciju modela na temelju detaljnih i pouzdanih subjektivnih procjena kvalitete slika. Ova baza podataka predstavlja značajan korak naprijed u razumijevanju i poboljšanju percepcije kvalitete slika generiranih umjetnom inteligencijom, te doprinosi razvoju naprednijih i učinkovitijih generativnih modela. Uzorci slika iz AGIQA-3K baze podataka prikazani na Slici 5.3.



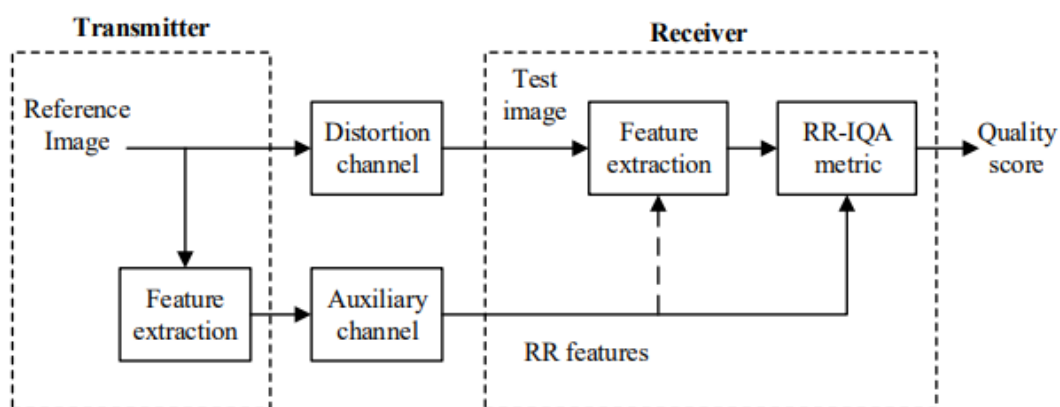
Slika 5.3 Uzorci slika iz AGIQA-3K baze podataka, gdje prvi do šesti redak prikazuju slike generirane pomoću modela (AttnGAN, DALLE2, GLIDE, Midjourney, Stable Diffusion i Stable Diffusion XL), dok stupci označavaju isti ulazni prompt [22].



Slika 5.4 (engl. artificial general intelligence, AGI), gdje su ključne riječi za generiranje označene u gornjem desnom kutu. (a) Tehnički problemi su niskorazinska izobličenja poput zamućenja i besmislenih objekata. (b) Artefakti umjetne inteligencije znače neočekivane artefakte kao što su nedostajući/višak udova. (c) Deepfake se odnosi na AGI slike koje HVS prepoznaje kao lažne zbog nerealnih proporcija, osvjetljenja, itd. (d) Estetski aspekti ukazuju na estetsku kvalitetu uključujući nedostatak detalja, neugodne objekte itd. [22].

6. Objektivna evaluacija kvalitete slike

Objektivna evaluacija kvalitete slike ključna je za brojne primjene u obradi slike, gdje je cilj metoda za procjenu kvalitete slike automatski ocijeniti kvalitetu slika u skladu s ljudskim prosudbama kvalitete [18]. Idealna metoda objektivne procjene kvalitete slike (IQA) trebala bi moći oponašati predviđanja kvalitete prosječnog ljudskog promatrača. Postoje tri glavne kategorije metoda objektivne procjene kvalitete slike, temeljene na dostupnosti referentne slike koja se smatra bez izobličenja i savršene kvalitete. Prva kategorija je procjena kvalitete slike s referentnom slikom (FR-IQA), gdje je dostupna nepromijenjena, savršena referentna slika. Ove metode uspoređuju testnu sliku s referentnom slikom kako bi izmjerile kvalitetu. Druga kategorija je procjena kvalitete slike s reduciranom referencom (RR-IQA), gdje referentna slika nije u potpunosti dostupna, već se umjesto toga koriste određene značajke referentne slike kao pomoćne informacije za procjenu kvalitete testne slike. Okvir RR-IQA sustava za procjenu kvalitete slike s reduciranom referencom prikazan je na Slici 6.1.



Slika 6.1 Okvir RR-IQA sustava za procjenu kvalitete slike s reduciranom referencom [18].

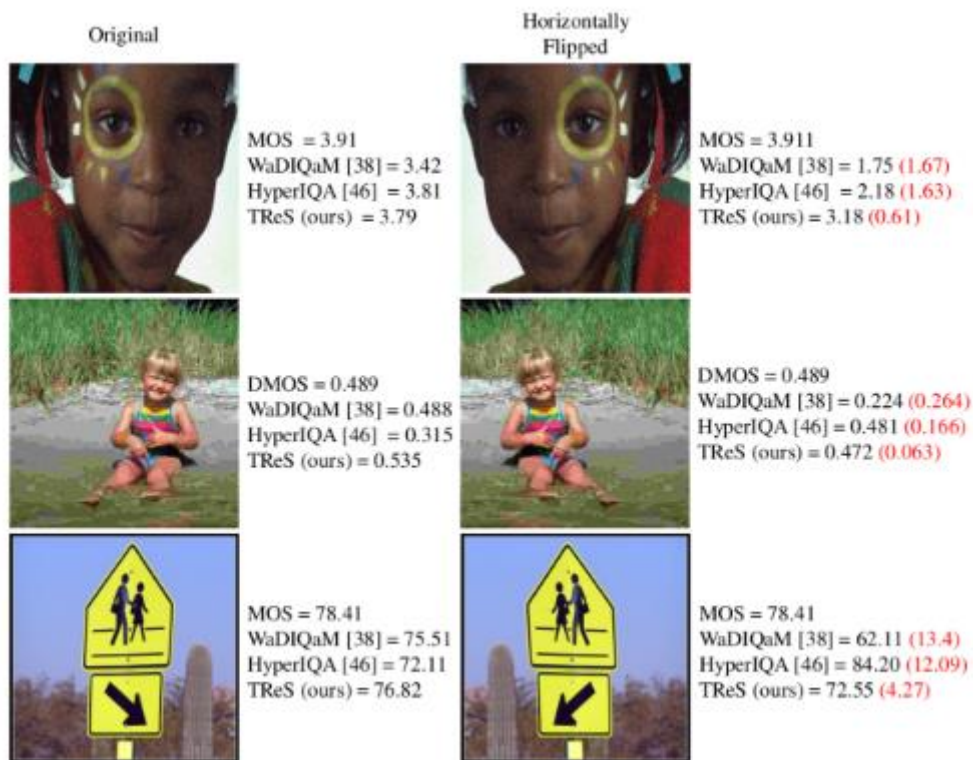
Treća kategorija je procjena kvalitete slike bez reference (NR-IQA), gdje referentna slika ili njezine značajke nisu dostupne za evaluaciju kvalitete. Ove metode su posebno korisne u stvarnim aplikacijama gdje referentna slika nije dostupna. Objektivne metode procjene kvalitete slike mogu se također kategorizirati prema svojoj primjeni. Opće namjene metode ne pretpostavljaju specifičnu vrstu izobličenja i korisne su u širokom rasponu aplikacija. S druge strane, metode specifične za primjenu dizajnirane su za specifične vrste izobličenja, poput algoritama za kompresiju slike. Različiti modeli FR-IQA uključuju mjere kao što su srednja kvadratna pogreška (engl. *mean squared error*, MSE), strukturni indeks sličnosti (engl. *structural similarity index measure*, SSIM), strukturni indeks sličnosti na više skala (engl. *multiscale structural similarity index measure*, MS-SSIM), vizualna informacijska vjernost (engl. *visual information fidelity*, VIF), najvidljivije izobličenje (engl. *maximum apparent distortion*, MAD) i indeks sličnosti značajki (engl. *feature similarity index*, FSIM). Ovi modeli su temeljito opisani i procijenjeni

prema njihovoj učinkovitosti i vremenu na različitim skupovima podataka subjektivne kvalitete. Razvoj metoda za procjenu kvalitete slike bez reference i s reduciranom referencom predstavlja poseban izazov, jer se kvaliteta procjenjuje isključivo na temelju testne slike ili uz ograničene informacije iz referentne slike. Ipak, ove metode su ključne za stvarne aplikacije gdje referentna slika često nije dostupna.

6.1. Evaluacija kvalitete slike bez reference

Evaluacija kvalitete slike bez reference (NR-IQA) nastoji procijeniti percepcijsku kvalitetu slike bez pristupa referentnoj slici koja se smatra savršenom. Ova metoda je izazovna i neriješena zbog nedostatka izvorne, neizobličene referentne slike. Cilj NR-IQA je osigurati procjenu kvalitete slike koja se podudara s ljudskim subjektivnim ocjenama, ali to je složen problem zbog različitih vrsta izobličenja koje slike mogu sadržavati. NR-IQA metode se mogu podijeliti u dvije glavne skupine: metode temeljene na specifičnim izobličenjima i opće namjene. Metode temeljene na izobličenjima fokusiraju se na procjenu kvalitete za određenu vrstu izobličenja, kao što su zamućenje ili izobličenja zbog kompresije. Međutim, ove metode imaju ograničenu primjenu u stvarnim scenarijima jer nije uvijek moguće unaprijed odrediti vrstu izobličenja. Stoga su razvijene metode opće namjene koje koriste značajke informativne za različite vrste izobličenja, čime se omogućava šira primjena. Tradicionalno, metode NR-IQA opće namjene fokusirale su se na procjenu kvalitete sintetički izobličenih slika, poput zamućenja, JPEG kompresije i Gaussovog šuma. Glavni izazovi kod ovih sintetički izobličenih skupova podataka su ograničena raznolikost sadržaja i izobličenja te nemogućnost ispitivanja složenih kombinacija izobličenja koje se često javljaju u stvarnim slikama. Nedavno su uvedeni skupovi podataka iz stvarnog svijeta, poput CLIVE, KonIQ-10K i LIVEFB, koji omogućuju bolje razumijevanje složenih izobličenja (npr. loši uvjeti osvjetljenja, ograničenja senzora, nesavršenosti objektiva, amaterske manipulacije) koja se često javljaju u stvarnim slikama. Postojeće NR-IQA metode temeljene na dubokom učenju uglavnom se oslanjaju na subjektivne ljudske ocjene (MOS/DMOS) i modeliraju zadatak predikcije kvalitete kao problem regresije ili klasifikacije. Ove metode često zanemaruju relativni poredak među slikama. Kako bi se to prevladalo, predložena je mjera TReS (engl. *transformers, relative ranking, and self-consistency*) koja uzima u obzir relativne udaljenosti među slikama unutar svakog skupa podataka i nameće relativni poredak među njima [23]. Također je predložena metoda koja koristi samo konzistentnost kao izvor samostalnog nadzora za poboljšanje robusnosti NR-IQA modela. Kombinirajući lokalne značajke koje generiraju CNN-i sa mehanizmom samopažnje transformer mreže, moguće je postići bolju procjenu kvalitete slike. Transformer modeli koriste se za modeliranje ne-lokalnih odnosa između značajki izvučenih s više skala različitih slojeva CNN-a. Ove značajke zatim se kombiniraju za predikciju konačne ocjene

kvalitete slike. NR-IQA predstavlja kritičan alat za različite aplikacije računalnog vida te za društvene i internetski prijenos u medijskoj industriji, gdje točna procjena kvalitete slike može imati veliki utjecaj na korisničko iskustvo. Ilustracija osjetljivosti NR-IQA modela na horizontalno okretanje nalazi se na Slici 6.2.



Slika 6.2 Ilustracija osjetljivosti NR-IQA modela na horizontalno okretanje. Na desnoj strani svake slike, prikazana je subjektivna ocjena kvalitete (MOS/DMOS) i predviđena ocjena kvalitete; crveni brojevi u zagradama pokazuju apsolutnu razliku između predikcija kada je slika okrenuta [23].

6.2. NIQE mjera kvalitete slike

NIQE mjera (engl. *natural image quality evaluator*) je model za procjenu kvalitete slike bez reference koji ne zahtijeva treniranje na ljudskim ocjenama ili prethodno poznavanje izobličenja slike [27]. Za razliku od drugih metoda, NIQE koristi mjerljive devijacije od statističkih pravilnosti opaženih u prirodnim slikama. Ove značajke se izvlače iz skupa prirodnih, neizobličenih slika, čime NIQE postaje potpuno neovisan o treniranju na izobličenim slikama ili ljudskim ocjenama (engl. *opinion-unaware*). NIQE se temelji na konstrukciji skupa značajki koje su osjetljive na kvalitetu slike i koje su izvedene iz jednostavnog, ali učinkovitog modela prirodnih scena. Ove značajke se zatim koriste za izračunavanje multivarijatnog Gaussovog (MVG) modela (engl. *multivariate Gaussian model*, MVG). Kvaliteta testirane slike izražava se kao udaljenost između MVG modela značajki kvalitete skupa prirodnih slika i MVG modela značajki izvučenih iz testirane slike. Jedna od ključnih prednosti NIQE metode je da može procijeniti kvalitetu slika bez

potrebe za specifičnim informacijama o izobličenjima ili treniranju na ljudskim ocjenama. Ovo čini NIQE vrlo praktičnim za upotrebu u stvarnim aplikacijama gdje izobličenja nisu unaprijed poznata, poput web stranica za prijenos fotografija ili različitih multimedijских platformi. Eksperimentalni rezultati pokazuju da NIQE postiže usporedive rezultate s najboljim metodama procjene kvalitete slika s najboljim metodama procjene kvalitete slika (u vrijeme ispitivanja) zahtijevaju treniranje na ljudskim ocjenama, te nadmašuje mnoge postojeće modele za procjenu kvalitete slike bez reference. Ova metoda pokazuje značajnu moć i široku primjenjivost u procjeni kvalitete slike, pružajući korisne alate za razne industrije koje se bave obradom i prijenosom vizualnih sadržaja.

6.3. PIQUE mjera kvalitete slike

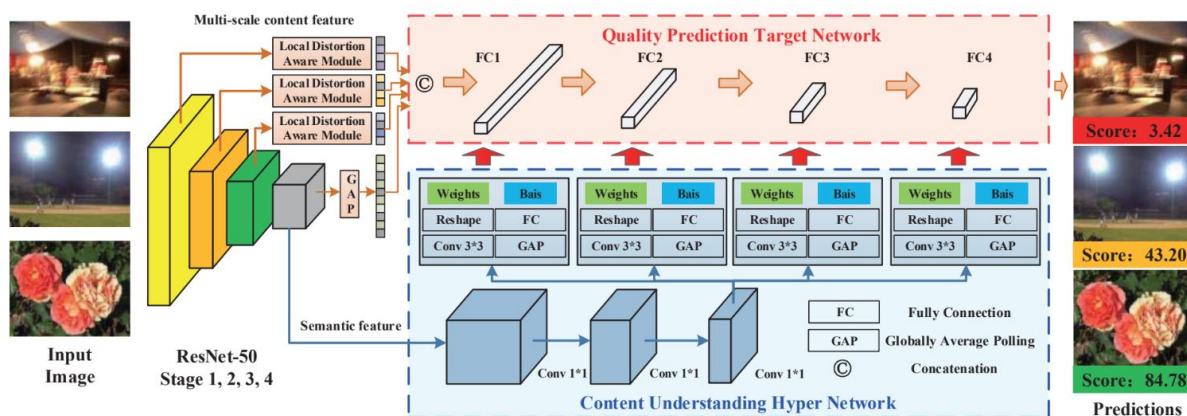
Kako bi se riješio izazov tradicionalnih metoda procjene kvalitete slike, razvijen je novi pristup nazvan PIQUE ili PIQE (engl. *perception-based image quality evaluator*), koji procjenjuje kvalitetu slika bez potrebe za podacima za treniranje ili informacijama o izobličenjima [26]. PIQE koristi lokalne značajke slike za predviđanje kvalitete, fokusirajući se na perceptualno značajna prostorna područja slike, imitirajući ljudsko ponašanje. PIQE je mjera koja nije trenirana subjektivnim ocjenama (engl. *opinion-unaware*), tj. kvantificira distorziju bez potrebe za bilo kakvim podacima za trening. PIQE može mjeriti kvalitetu slika s proizvoljnim izobličenjem i u većini slučajeva radi slično kao NIQE [27]. PIQE procjenjuje izobličenje po blokovima i mjeri lokalnu varijancu vizualno izobličenih blokova kako bi se izračunao rezultat kvalitete. PIQE također generira detaljnu prostornu kartu kvalitete, koja je korisna za različite primjene poput ekstrakcije značajki, detekcije objekata i kompresije. Eksperimentalni rezultati pokazuju da PIQE nadmašuje postojeće metode na izazovnim autentičnim skupovima podataka, dok postiže konkurentne rezultate na sintetičkim skupovima podataka, iako nije izričito dizajniran za sintetičke zadatke. PIQE je brz i može se koristiti u stvarnim aplikacijama s niskim zahtjevima za računalnu složenost, što ga čini pogodnim za stvarnu upotrebu u raznim industrijama, uključujući društvene mreže i internetski prijenos u medijskoj industriji. Ovaj pristup značajno doprinosi području NR-IQA, omogućujući bolje razumijevanje i procjenu kvalitete slika bez potrebe za referentnim slikama. JP2K izobličena slika iz LIVE baze podataka s dodatnim AWGN šumom i PIQE prostornom maskom kvalitete prikazana na Slici 6.4.



Slika 6.3 JP2K izobličena slika iz LIVE baze podataka s dodatnim AWGN šumom i PIQE prostornom maskom kvalitete. Crveno, žuto i zeleno ispunjeni blokovi označavaju blokove koji zadovoljavaju NDC, NC i uniforme kriterije [27].

6.4. HyperIQA mjera kvalitete slike

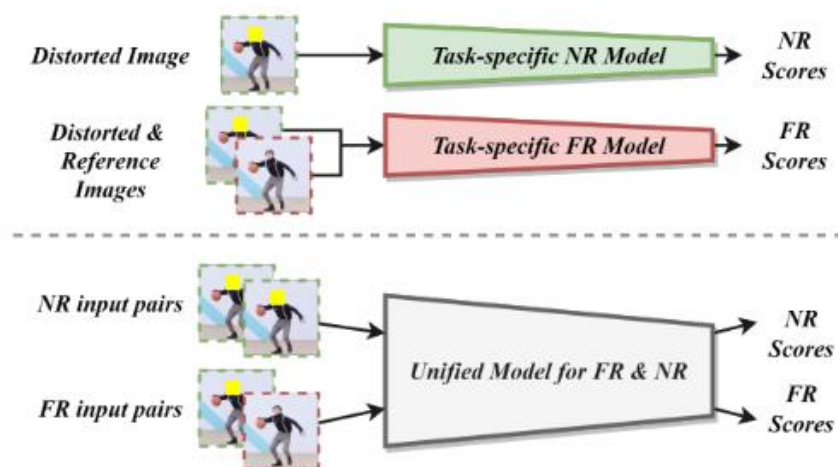
Metode bez reference nastoje procijeniti kvalitetu slike na temelju njenih karakteristika bez usporedbe s originalnom slikom. Kako bi se riješio ovaj izazov, istraživači su razvili samoadaptivnu hiper-mrežu HyperNet koja procjenjuje kvalitetu slika u stvarnim uvjetima [25]. Ova mreža dijeli postupak procjene kvalitete na tri faze: razumijevanje sadržaja, učenje perceptivnih pravila i predikciju kvalitete. Nakon što izvadi semantičke značajke slike, hiper-mreža adaptivno uspostavlja pravila percepcije, koja se potom koriste za predikciju kvalitete slike. Ovaj pristup omogućava mreži da se prilagodi raznovrsnim slikama i pruža točnije procjene kvalitete koje su u skladu s ljudskom percepcijom. Eksperimentalni rezultati pokazuju da ovaj pristup ne samo da nadmašuje postojeće metode na izazovnim autentičnim bazama podataka, već postiže i konkurentne rezultate na sintetičkim bazama podataka, iako nije izričito dizajnirana za sintetičke zadatke. Time se dokazuje moć i široku primjenjivost predložene metode u procjeni kvalitete slike bez reference. Proces proračunavanja predložene mjere HyperIQA prikazan je na slici 6.3.



Slika 6.4 Proces proračunavanja predložene mjere HyperIQA [25].

6.5. UNIQA mjera kvalitete slike

Digitalne slike su prisutne posvuda, s primjenama koje sežu od potrošačke fotografije do medicinskog slikanja i satelitskog daljinskog osjetljivosti. Važnost točnog vrednovanja kvalitete slika postaje sve važnija u mnogim područjima. Iako je ljudski vizualni sustav učinkovit u prepoznavanju visokokvalitetnih slika, procjena kvalitete slike putem ljudskih radnika često je radno intenzivna i u većini slučajeva neizvediva. Stoga se ocjena kvalitete slike (IQA) bavi razvojem objektivnih metrika koje mogu predviđati kvalitetu slike onako kako je doživljavaju ljudi. Ovisno o prisutnosti visokokvalitetnih referentnih slika, zadaci IQA mogu se podijeliti u dvije kategorije: ocjena bez referenci (NR) i ocjena s punim referencama (FR). U doba dubokog učenja, osnovna ideja većine metoda FR-IQA je procijeniti kvalitetu ocjenom niskokvalitetnih područja putem razlike između referentnih i izobličenih slika. Za NR-IQA, uobičajene prakse uključuju rangiranje, usporedbu značajki, generiranje izvornih slika pomoću generativnih suparničkih mreža (GAN) i primjenu transformer neuronskih mreža ili višeslojnih CNN-ova za bolje prostorno modeliranje izobličenja. Unatoč sličnostima, mreže su obično dizajnirane za obavljanje samo jednog zadatka, što ograničava njihovu primjenu i može uzrokovati nesukladnost performansi pri prelasku s FR-IQA na NR-IQA. Ilustracija glavne svrhe predložene mreže gdje se različiti parovi slika mogu unositi u mrežu kako bi se dobili FR/NR IQA rezultati koristeći istu arhitekturu nalaze se na Slici 6.5. U ovom radu predložen je jedinstveni okvir za oba zadatka, koji uključuje modul hijerarhijske samopažnje (engl. *hierarchical self-attention*, HSA) za prilagodbu FR-IQA i NR-IQA ulaze, te modul križne pažnje s unakrsnim skalama (engl. *cross-scale cross-attention*, CSCA) za proučavanje korelacija između izobličenja na plitkim i dubokim razinama. Predložena mreža pokazuje učinkovitost na sintetički izobličenim i autentično izobličenim skupovima podataka, nadmašujući relevantne metode na obje vrste skupova podataka [24].

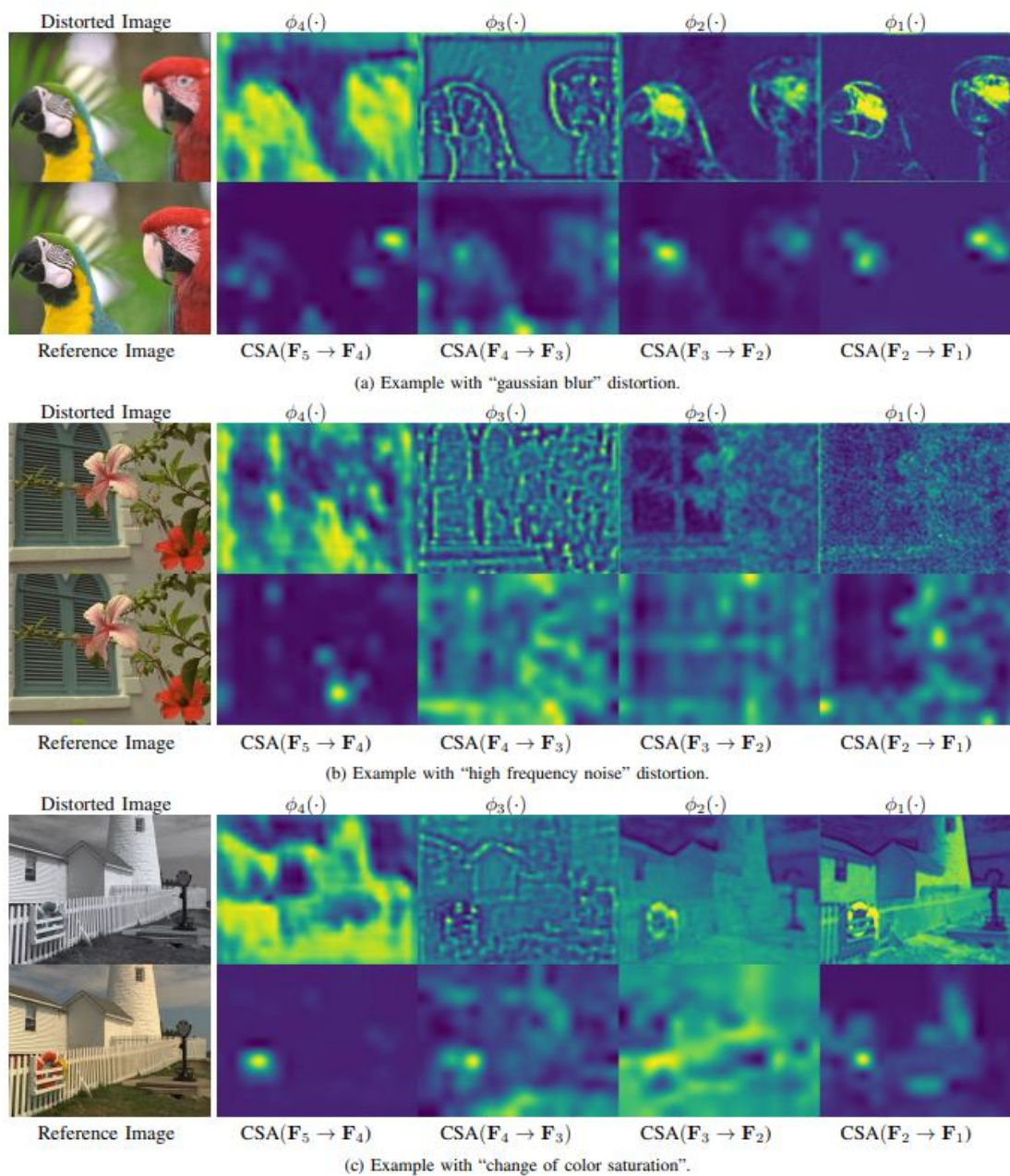


Slika 6.5 Ilustracija glavne svrhe predložene mreže (donji dio slike), gdje se različiti parovi slika mogu unositi u mrežu kako bi se dobili FR/NR IQA rezultati koristeći istu arhitekturu. Tradicionalni

način izračunavanja kvalitete slike prikazan je kao gornji dio slike radi jednostavnije usporedbe [24].

6.6. TOPIQ mjera kvalitete slike

TOPIQ mjera (engl. *top-down IQA*) je inovativan pristup koji koristi semantičke informacije visoke razine kako bi usmjerio mrežu za procjenu kvalitete slike na fokusiranje na lokalna izobličenja od semantičke važnosti [28]. Ovaj pristup uključuje razvoj heurističke mreže pažnje grubo-do-fino (CFANet), koja koristi značajke na više skala i progresivno prenosi višerazinsku semantičku informaciju prema nižim razinama na način odozgo do dolje. Ključna komponenta mjere TOPIQ je predloženi mehanizam pažnje između skala (engl. *cross-scale attention*, CSA), koji izračunava mape pažnje za niže razine značajki vođene višim razinama značajki. CFANet se može koristiti za FR-IQA i NR-IQA procjenu kvalitete slike. U CFANet modelu, ResNet50 mreža služi kao osnova modela (engl. *backbone*). CFANet postiže bolje ili slične performanse na većini javno dostupnih baza slika za ispitivanje FR-IQA i NR-IQA modela u usporedbi s najmodernijim metodama temeljenim na vizualnim transformer mrežama (ViT), pri čemu je znatno učinkovitiji. TOPIQ pristup naglašava korištenje dubokih značajki na više skala za IQA. Eksperimentalni rezultati pokazuju da CFANet ne samo da nadmašuje postojeće metode na izazovnim autentičnim bazama podataka, već također postiže i konkurentne rezultate na sintetičkim bazama podataka, iako nije izričito dizajniran za sintetičke zadatke. Ovi rezultati potvrđuju učinkovitost i široku primjenjivost predložene metode TOPIQ u procjeni kvalitete slike bez reference. Vizualizacija pažnje s različitim vrstama izobličenja iz TID2013 skupa podataka prikazana je na Slici 6.6.



Slika 6.6 Vizualizacija pažnje s različitim vrstama izobličenja iz TID2013 skupa podataka. Prvi redak: Maske modula ograničenog lokalnog udruživanja (engl. gated local pooling, GLP) na 4 razine; Drugi redak: CSA težine [28].

7. AI alati za generiranje slika

Umjetna inteligencija (UI), ili kako se često naziva prema engleskoj skraćenici AI (Artificial Intelligence), dio je informatike koji se fokusira na razvijanje sposobnosti računala da obavljaju zadatke koji zahtijevaju određeni stupanj inteligencije. Također, odnosi se na svojstvo sustava koji, iako neživi, pokazuju inteligentno ponašanje. Razvoj umjetne inteligencije zasniva se na konceptu strojnog učenja, posebno neuronskih mreža, što omogućuje ovim sustavima da usvoje ogromne količine informacija, komuniciraju s ljudima (preko prirodnog jezika) ili s drugim uređajima, uče iz iskustva, donose odluke, prilagođavaju se situacijama, planiraju i slično. Ipak, današnji inteligentni sustavi još uvijek su usmjereni na specifična područja, dok je čovjek sposoban obavljati mnogobrojne zadatke istovremeno. Umjetna inteligencija dijeli se na "slabu" i "jaku" ovisno o stupnju razvijenosti. Svi današnji sustavi spadaju u kategoriju slabe AI, gdje posjeduju samo neka svojstva inteligencije. S druge strane, jaka AI, koja još nije postignuta, trebala bi imati razinu inteligencije sličnu ljudskoj (opća umjetna inteligencija) ili ju nadmašiti (superinteligencija), uključujući svijest. Područje primjene umjetne inteligencije je široko i raznoliko. Danas se koristi u strojnom prevođenju, ekspertima sustavima, robotici (posebno u kretanju i motorici), računalnim igrama i simulacijama, generiranju teksta, chatbotovima, stvaranju glazbe i slika, računalnom vidu (prepoznavanju objekata, analizi scena), pretraživanju podataka, automatskom programiranju i još mnogo toga. Ideje o stvaranju inteligentnih strojeva mogu se pratiti kroz povijesne mitove i književnost, poput priče o Golemu. Ipak, razvoj umjetne inteligencije započeo je s pojavom elektroničkih računala u drugoj polovici 20. stoljeća. Mnoge metode preuzete su iz različitih znanstvenih područja, uključujući kognitivne znanosti, matematiku, logiku, filozofiju, psihologiju i lingvistiku, koje se bave razumijevanjem ljudskog mišljenja. Danas, programi poput DALL-E i Midjourney omogućuju generiranje slika na temelju tekstualnih uputa, dok ChatGPT, veliki jezični model, pruža razne funkcije poput pisanja tekstova, programiranja i analize. Napredni sustavi za pomoć vozačima, poput ADAS-a, koriste računalni vid za povećanje sigurnosti na cestama. Iako umjetna inteligencija ima ključnu ulogu u svakodnevnoj tehnologiji, poput internetskih pretraživača, njezin razvoj pokreće brojne rasprave. Pitanja o autorskim pravima, privatnosti podataka, sigurnosti i zloupotrebi, kao što su deepfake tehnologije, postaju sve relevantnija. Također, izazovi poput "crne kutije", gdje je teško razumjeti kako algoritmi donose odluke, otvaraju dodatna etička i pravna pitanja u vezi s odgovornošću i primjenom ovih tehnologija [32].

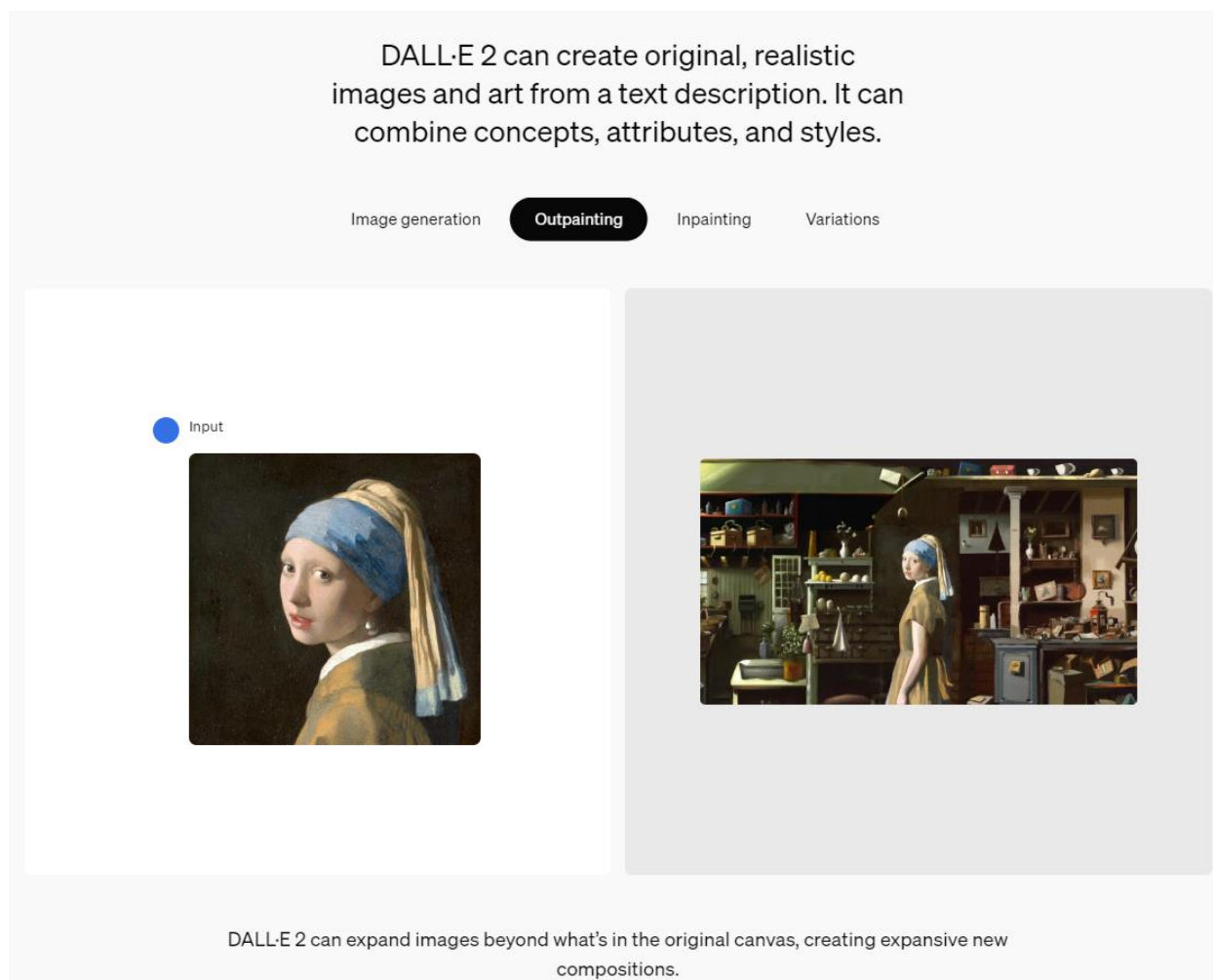
7.1. OpenAI i ChatGPT

OpenAI je organizacija posvećena razvoju napredne umjetne inteligencije, a jedan od njezinih najpoznatijih proizvoda je ChatGPT, sofisticirani jezični model razvijen za generiranje prirodnog jezika. ChatGPT je posebno dizajniran za obavljanje raznih zadataka u obradi prirodnog jezika, poput generiranja teksta, odgovaranja na pitanja i vođenja razgovora. Razvijen je na temelju GPT (Generative Pretrained Transformer) arhitekture, koja omogućuje razumijevanje konteksta i generiranje relevantnih odgovora. Proces treniranja ChatGPT-a odvija se u dvije faze: prvo je pre-treniran na ogromnim količinama tekstualnih podataka kako bi stekao opće razumijevanje jezika, a zatim je podešen za specifične zadatke poput dovršavanja teksta ili generiranja dijaloga. Model je iznimno sposoban zahvaljujući svojoj arhitekturi koja uključuje "self-attention" mehanizam, što mu omogućava fokusiranje na različite dijelove teksta i stvaranje koherentnih odgovora. Također koristi kodiranje koje omogućava razbijanje složenih ili rijetkih riječi na manje dijelove, čime uklanja nepotrebne informacije tijekom obrade podataka. ChatGPT je javnosti postao dostupan krajem 2022. godine i odmah je privukao veliku pažnju zbog svojih mogućnosti. Korišten je u različitim područjima, uključujući obrazovanje, zdravstvo, kreativno pisanje i marketing. Njegova sposobnost generiranja realističnog i smislenog teksta omogućava prirodniju interakciju između ljudi i strojeva. Uz sve svoje prednosti, ChatGPT ipak ima ograničenja, uključujući povremene pristranosti u podacima na temelju kojih je treniran te poteškoće u rješavanju složenijih logičkih zadataka [33].

7.2. DALL-E

DALL-E je sustav koji je razvio OpenAI, a omogućava generiranje originalnih slika na temelju unosa teksta. Druga verzija sustava, DALL-E 2, još je impresivnija u svojoj sposobnosti stvaranja sintetičkih slika koje odgovaraju opisu unesenom u tekstualnom obliku. Koristi naprednu tehnologiju obrade prirodnog jezika i računalnog vida kako bi "razumio" tekst i pretvorio ga u vizualni prikaz. Sustav DALL-E 2 pokazuje izuzetne rezultate u primjeni različitih umjetničkih stilova, poput crtanja, fotografije, impresionističkog slikarstva i drugih, vjerno prenoseći specifičnosti svakog stila. Uz to, slike generirane u realističnim stilovima gotovo uvijek djeluju fizički uvjerljivo. Ipak, sustav ima problema s razumijevanjem složenijih zahtjeva koji uključuju odnose između objekata ili apstraktnije pojmove. Iako je DALL-E 2 sposoban pratiti upute koje se tiču broja i osnovnih karakteristika objekata, suočava se s poteškoćama kada je potrebno razumjeti složene prostorne odnose ili apstraktne koncepte. Primjeri su neprecizna pozicioniranja objekata u složenim scenama ili nerazumijevanje zadataka vezanih uz brojanje i logičke relacije.

Jedna od impresivnijih karakteristika DALL-E 2 je njegova sposobnost stvaranja slika visoke kvalitete koje se mogu koristiti u različite svrhe, od umjetničkih do komercijalnih. Međutim, kod primjene u situacijama koje zahtijevaju visok stupanj pouzdanosti, poput sigurnosno osjetljivih aplikacija, DALL-E 2 još uvijek nije na razini koja bi omogućila potpuno povjerenje. Uz sve pozitivne strane, DALL-E 2 ima određene nedostatke, uključujući ograničenja u razumijevanju negacije, brojanja, kao i problema s povezivanjem pojmova unutar složenijih tekstualnih opisa. U određenim slučajevima, sustav ne uspijeva ispravno interpretirati apstraktne zadatke, kao što su složeni odnosi između objekata [34]. Slika 7.1. prikazuje web stranicu na kojoj se nalazi pregled osnovnih funkcionalnosti alata DALL-E 2.

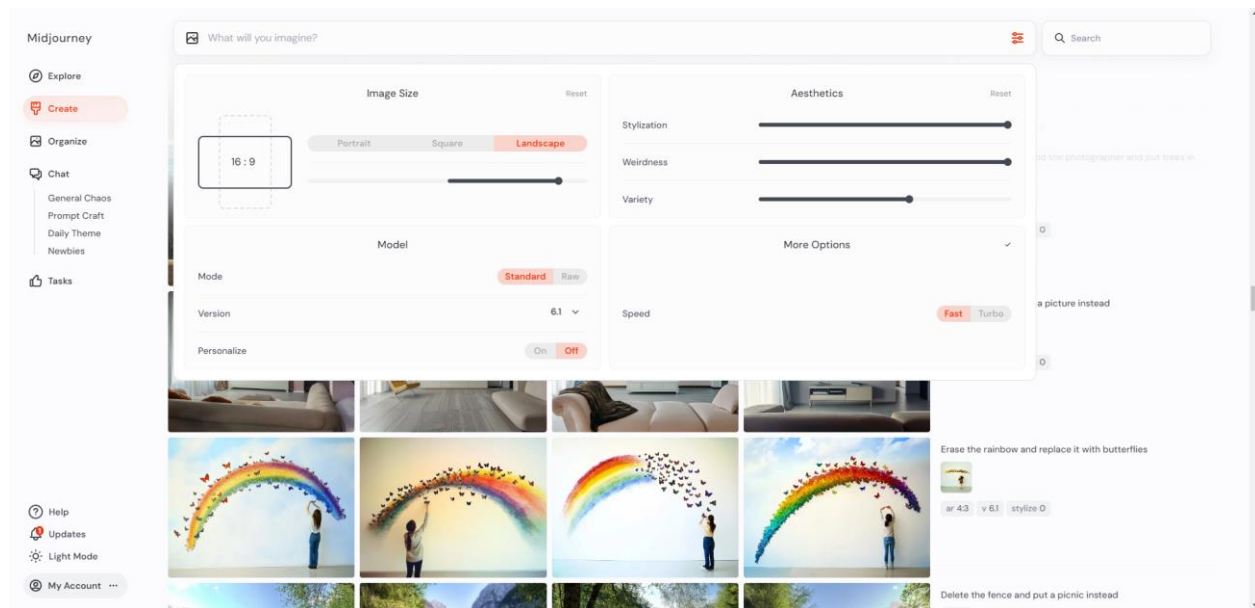


Slika 7.1 Prikaz funkcionalnosti alata DALL-E 2 [35].

7.3. Midjourney

Midjourney je program umjetne inteligencije koji kreira slike na temelju opisa unesenih putem prirodnog jezika, slično kao DALL-E od OpenAI i Stable Diffusion od Stability AI. Program je razvijen od strane neovisnog istraživačkog laboratorija Midjourney, Inc., smještenog u San Franciscu. Alat je dostupan u otvorenoj beta verziji od srpnja 2022. godine, a korisnici mogu

stvarati slike koristeći Discord bot komande ili putem službene web stranice. Midjourney je postao popularan među umjetnicima, dizajnerima i oglašivačima jer omogućava brzo kreiranje vizualnog sadržaja bez potrebe za dugotrajnim crtanjem ili slikanjem. Prikaz funkcionalnosti alata Midjourney nalazi se na Slici 7.2. Kroz razne verzije svog algoritma, Midjourney je napredovao, pružajući poboljšane funkcije i sposobnosti. Na primjer, verzija 5.2 omogućila je opciju "zumiranja" postojećih slika i bolju interpretaciju estetskih parametara, dok je verzija 6, izdana krajem 2023., donijela poboljšanja u prikazu teksta i doslovnoj interpretaciji unosa [36].



Slika 7.2 Prikaz funkcionalnosti alata Midjourney [38].

7.4. Autorska prava generiranog sadržaja

Autorska prava u kontekstu slika generiranih umjetnom inteligencijom (AI), poput onih koje stvaraju sustavi kao što su DALL-E, Midjourney i Stable Diffusion, predstavljaju složeno pitanje. Ključno pitanje je može li generativna umjetna inteligencija stvoriti djelo koje potencijalno krši postojeća autorska prava te hoće li obrana temeljem "dozvoljene uporabe" biti valjana. Generativni AI sustavi stvaraju slike koristeći velike količine podataka, često prikupljenih s interneta, uključujući slike zaštićene autorskim pravima. To otvara pitanje autorskih prava, jer se može tvrditi da su te slike korištene bez pristanka autora. Međutim, kako AI sustavi funkcioniraju tako da stvaraju nove slike temeljene na obrazovnim modelima, a ne jednostavnim kopiranjem postojećih slika, postavlja se pitanje je li riječ o kršenju autorskih prava ili stvaranju potpuno novih djela. Također, uloga krajnjeg korisnika u stvaranju slika pomoću generativnih AI sustava ključna je u raspravi o autorskim pravima. Korisnici su ti koji osmišljavaju i unose promptove (tekstualne upute) koje pokreću proces generiranja slika. Na temelju tih uputa, AI sustav generira sliku, što znači da korisnik djelomično sudjeluje u procesu stvaranja, iako AI sustav obavlja tehnički dio

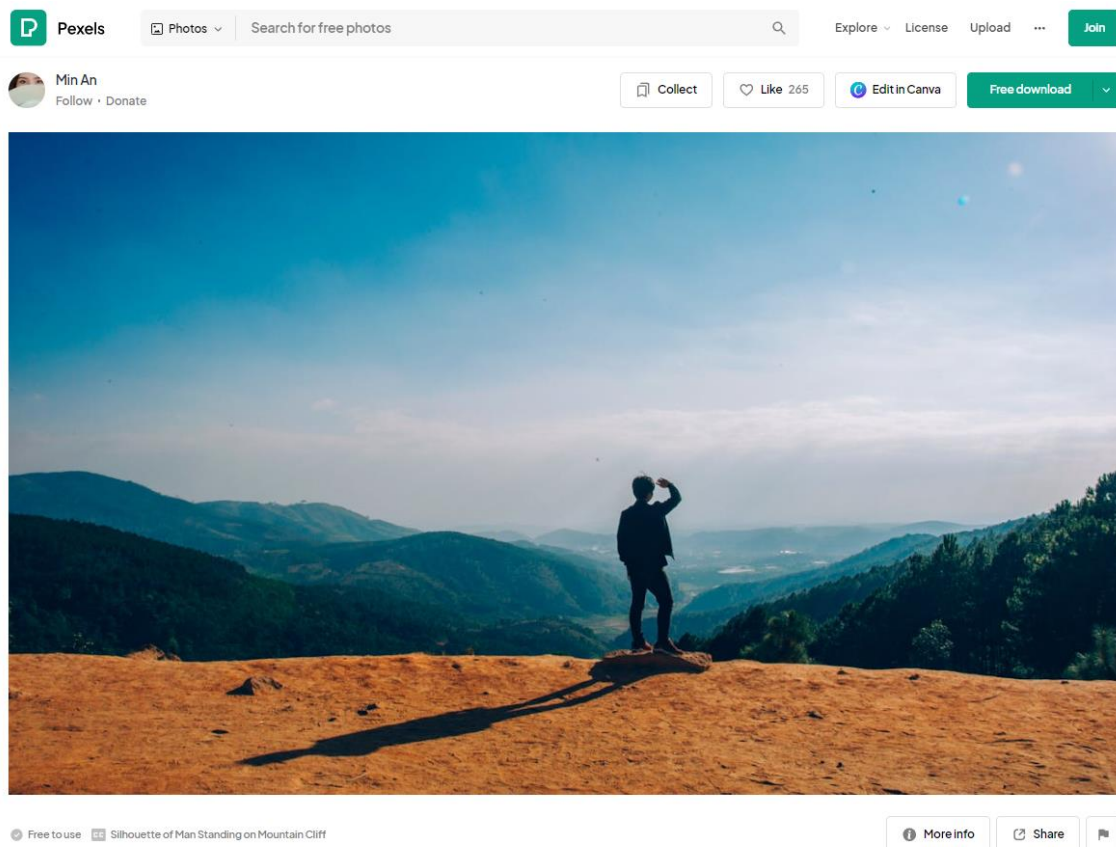
posla. Pravo na "dozvoljenu uporabu" može se primijeniti ako se generativni AI sustav koristi na način koji transformira izvorni materijal u nešto novo i drugačije. Prema zakonu o autorskim pravima, "dozvoljena uporaba" omogućuje korištenje zaštićenih djela bez dopuštenja, pod određenim uvjetima, poput kritike, komentara, obrazovanja ili istraživanja. No, sudovi će morati odrediti je li generativni AI stvarno transformativan ili samo ponovno koristi zaštićene elemente na štetu izvornih autora [37]. Jedna od poznatih kontroverzi vezanih uz Midjourney bila je pobjeda slike "Théâtre D'opéra Spatial", generirane putem Midjourney-a, na natjecanju digitalne umjetnosti, što je izazvalo nezadovoljstvo kod tradicionalnih umjetnika. Dodatno, korištenje AI za generiranje ilustracija, kao što je to bio slučaj s dječjom knjigom "Alice and Sparkle", izazvalo je kritike zbog korištenja stilova umjetnika bez njihove dozvole. Midjourney je također izazvao pozornost zbog kontroverznih AI-generiranih fotografija, poput lažne slike Pape Franje u puf jakni, što je pokazalo visok stupanj realizma u AI-generiranim slikama. Midjourney koristi sofisticirani sustav moderacije sadržaja koji je u 2023. godini poboljšao kako bi omogućio suptilniju analizu korisničkih unosa i spriječio generiranje kontroverznog ili uvredljivog sadržaja. Unatoč svojim prednostima i inovacijama, Midjourney se suočava s pravnim izazovima, uključujući tužbe zbog navodnog kršenja autorskih prava umjetnika čija su djela korištena u treniranju modela bez njihovog pristanka [36]. Trenutačna pravna situacija nije u potpunosti jasna, a pravni sustavi diljem svijeta tek počinju rješavati te izazove. Već su pokrenute brojne tužbe protiv tvrtki koje razvijaju generativne AI sustave, a ključne presude u tim slučajevima odredit će kako će se autorska prava primjenjivati u kontekstu umjetnosti generirane umjetnom inteligencijom.

8. Razvoj istraživanja

U praktičnom dijelu rada primijenit će se tehnike računalnog vida opisane u teorijskom dijelu, s posebnim naglaskom na testiranje sustava za popunjavanje slika i procjenu njihove kvalitete koristeći modele dubokog učenja. Sudjelovanje u istraživanju bilo je potpuno anonimno, a jedine prikupljene informacije o sudionicima bile su dob i spol. Sudionici su u bilo kojem trenutku mogli prekinuti sudjelovanje pritiskom na gumb X, koji je bio prisutan na svakoj stranici. Zadatak je bio ocijeniti 60 fotografija koje su generirali dva sustava temeljena na umjetnoj inteligenciji; DALL-E i Midjourney. Ovi sustavi stvaraju slike na temelju zadanih uputa i opisa. Oba sustava radila su s istim detaljnim opisima kako bi se testiralo njihovu sposobnost generiranja sadržaja prema zadanim tekstualnim uputama te usporedilo koliko vjerno generirane slike odgovaraju originalu. Za svaki sustav korišteno je 15 opisa u dva različita skupa, što je ukupno rezultiralo s 60 slika. Sudionici su ocjenjivali tri aspekta svake slike: opći dojam slike (vizualnu kvalitetu), autentičnost (koliko generirana slika odgovara stvarnoj fotografiji), te korespondenciju s tekstom (poklapanje teksta naredbe sa slikom). Koristio se klizač s rasponom od 1 do 5, pri čemu je 1 predstavljalo lošu, a 5 izvrsnu ocjenu. Prije procjene tih 60 slika, na testnoj stranici sudionicima su prikazane originalne slike uz jednu generiranu sliku na koju je bila primijenjena naredba proširi sliku i druga generirana slika sa naredbom za brisanje i zamjenu. Testne slike služile su za upoznavanje s metodom ocjenjivanja, no njezine ocjene nisu uzete u obzir. Kao što je ranije napomenuto, svaki sudionik mogao je odustati pritiskom na gumb X u gornjem desnom kutu, a do tada prikupljeni podaci nisu uključeni u završnu analizu.

8.1. Priprema sadržaja za istraživanje

Prije početka samog istraživanja trebalo se odlučiti koji sadržaj će se provući kroz umjetnu inteligenciju. Uzeto je 15 slika sa stranice pexels.com, slike su bile u potpunosti različitih sadržaja kako bi se što bolje prikazala raznolikost u generiranom sadržaju. Izabrana je horizontalna orijentacija radi skladnosti fotografija. Još jedan razlog odabira fotografija sa pexels.com je što su fotografije besplatne za preuzimanje, a kvaliteta samih slika je vrlo dobra. Izabrano je petnaest fotografija kako bi se što bolje moglo uočiti koji alat je pouzdaniji. Na Slici 8.1 se vidi jedna od fotografija korištenih u istraživanju, te sučelje web stranice Pexels.com.



Slika 8.1 Sučelje web stranice Pexels.com [39].

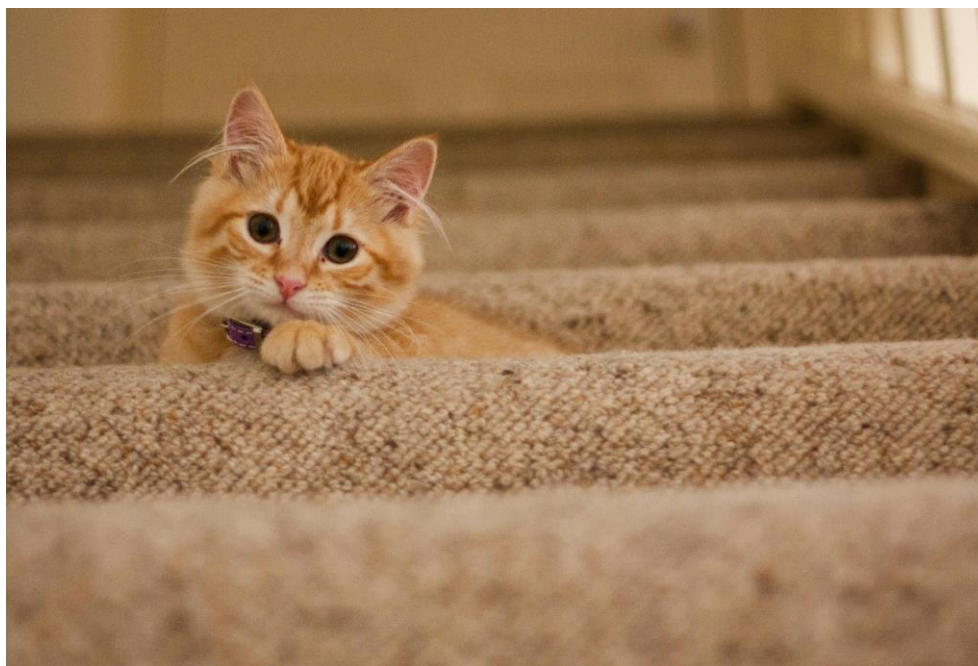
Nakon odabira fotografija uslijedilo je istraživanje koji od alata za generiranje slika bi odgovarao obzirom na osmišljen način ispitivanja. Neki od ponuđenih alata bili su DALL-E, Midjourney, Adobe Photoshop, SDXL 1.0. Obzirom na cijenu i funkcionalnosti, alati koji su najviše odgovarali bili su DALL-E i Midjourney, iz razloga što im je cijena bila pristupačna, a nije trebalo dodatno instalirati dodatne programe, jer se koriste preko web stranice.

Odabirom fotografija moglo se prijeći na odabir glavnih elemenata ispitivanja kvalitete slike, a to su naredbe koje su se zadale alatima umjetne inteligencije. Obzirom da se odlučilo za DALL-E i Midjourney, gledalo se da oba alata podržavaju naredbe proširi sliku (engl. *Expand Image*) i brisanje i zamjena (engl. *Erase and replace*). To su funkcije koje se koriste u alatima umjetne inteligencije za uređivanje i manipulaciju slika. Proširivanje slike omogućuje proširenje granica postojeće slike dodavanjem novih dijelova koji odgovaraju okolnom sadržaju slike. AI alat analizira postojeću sliku i stvara nove dijelove koji se prirodno uklapaju u cjelokupnu kompoziciju. Ova tehnika se često koristi za nadopunjavanje ili proširenje okvira slike, što je poznato i kao nadopunjavanje izvan slike (engl. *outpainting*). Brisanje i zamjena (engl. *Erase and replace*) omogućuje korisniku da ukloni dio slike (obično označen mišem ili nekom vrstom alata za odabir) i zatim zamijeni taj izbrisan dio nečim novim. AI analizira sliku i pomoću tekstualnog opisa može dodati novi sadržaj na mjesto izbrisanog dijela, dok se pokušava estetski i logički uklopiti s ostatkom slike.



8.2. Generiranje slika

Prilikom generiranja slika u oba alata primijećene su određene razlike. DALL-E je, na primjer, generirao dvije završne slike, dok je Midjourney ponudio četiri. Smatram da su četiri slike previše jer mi je taj proces oduzeo dosta vremena pri odlučivanju koju sliku koristiti za istraživanje. U DALL-E-u nije bilo dostupnih parametara za podešavanje prije generiranja slika, dok je Midjourney nudio niz opcija: veličina slike (portret, kvadrat, pejzaž), model (standardni ili sirovi način rada), verzije (od 1 do 6.1), prilagodbu (uključeno/isključeno), estetiku (stilizacija, čudnost, raznolikost na ljestvici od 0 do 1000), kao i dodatne opcije poput brzine (brzo ili turbo). Kako bi rezultati ostali dosljedni s obzirom na to da u DALL-E-u nisu postojale ove opcije, nisu korištena nikakva stiliziranja u Midjourney-u. DALL-E je nudio jednostavnije opcije, uključujući nekoliko stilova u kojima se može generirati slika, kao što su Art Nouveau, grafiti, ručno crtano, skulpture, ulje na platnu, te izbor omjera slike (kvadrat, široki zaslon, okomito).

Prve dvije ključne naredbe bile su "proširi sliku" (engl. expand) i "brisanje i zamjena" (engl. erase and replace). Zanimljivo je da DALL-E na osnovnom plaćenom modelu pokazuje latentnost nakon nekoliko uzastopnih generiranja slika. Nakon pet generiranih fotografija, pojavila se poruka o potrebi čekanja nekoliko minuta, zbog čega je proces usporen. Nasuprot tome, Midjourney ima prednost jer omogućuje nesmetano generiranje, čak i dvije slike odjednom, pri čemu je vrijeme generiranja svega nekoliko sekundi, dok DALL-E može trajati nekoliko minuta i trenutno generirati samo jednu sliku. Na Slici 8.2 vidi se originalna fotografija koja je preuzeta sa stranice Pexels.com te je korištena u istraživanju, zatim je ta slika stavljena u alate prikazane na Tablici 8.1.



Slika 8.2 Originalna fotografija korištena za generiranje naredbe prikazane u Tablici 8.1 [39].

Naredba	AI alat	
<p>Naredba brisanje i zamjena: „Izbriši mačku i stavi psa umjesto nje.“</p>	DALL-E	Midjourney
		

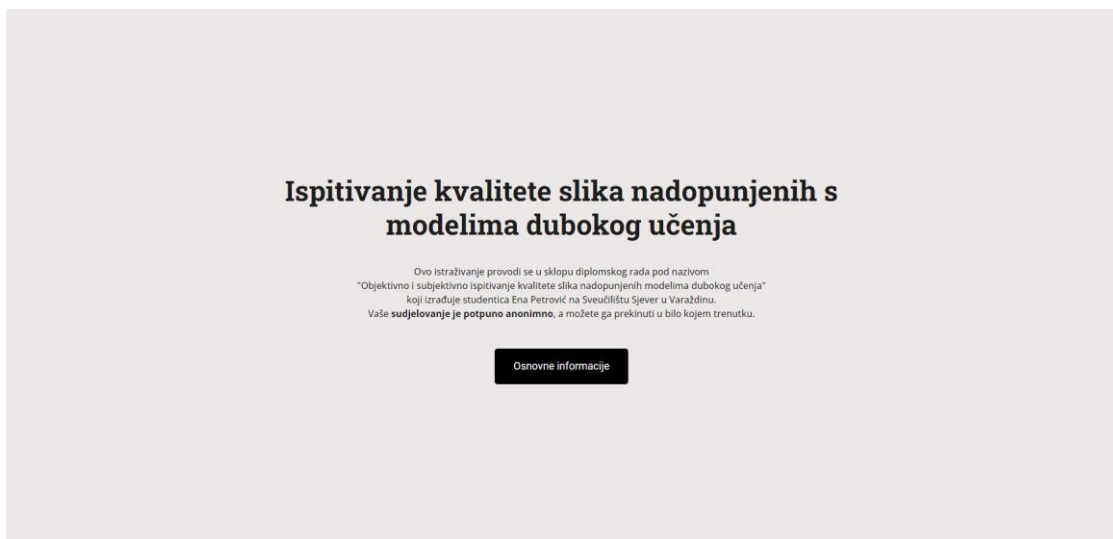
Tablica 8.1 Prikaz rezultata naredbe "brisanje i zamjena" u DALL-E i Midjourney alatima [izvor autor].

8.3. Dizajn i kreiranje web stranice

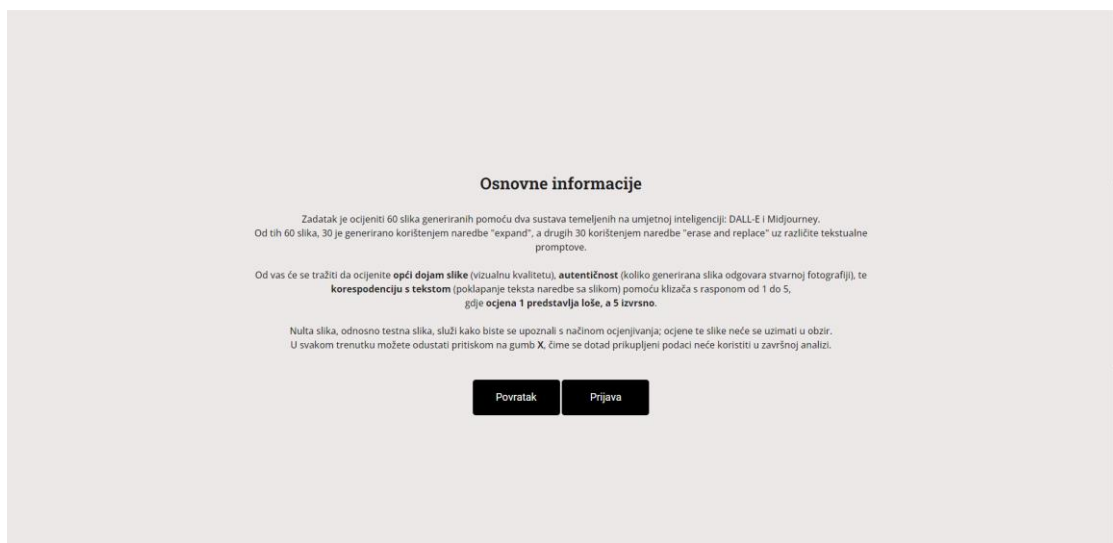
Za izradu web stranice korišten je PHP, koji je skriptni jezik namijenjen za rad na strani servera. PHP omogućuje generiranje dinamičkih stranica, interakciju s korisnicima i upravljanje podacima, a u ovom istraživanju koristi se za spremanje odgovora ispitanika u .txt datoteku pod nazivom "odgovori_ispitanika.txt". Ovaj jezik pomaže u kreiranju dinamičkog sadržaja i osigurava funkcionalnost stranice kroz komunikaciju sa serverom. Za stilizaciju web stranice korišten je CSS, koji omogućuje prilagodbu vizualnog izgleda elemenata na stranici. Kroz CSS definirane su boje, fontovi, raspored elemenata i cjelokupna estetika stranice. U ovom projektu korišteni su i fontovi "Roboto Slab" i "Open Sans", preuzeti putem web stranice Google Fonts, kako bi tekst bio vizualno atraktivniji i čitljiviji. HTML je bio ključan za strukturiranje osnovnih elemenata na stranici, omogućujući definiranje sadržaja poput tekstova, obrazaca i drugih ključnih komponenti. Za dodavanje interaktivnosti koristio se JavaScript, koji omogućava dinamično mijenjanje sadržaja stranice bez potrebe za ponovnim učitavanjem. JavaScript je, uz jQuery, popularnu biblioteku koja pojednostavljuje rad s HTML dokumentima, korišten kako bi se olakšala manipulacija stranicom i upravljanje događajima. Kako bi stranica mogla funkcionirati lokalno, korišten je XAMPP, koji omogućava postavljanje lokalnog servera i olakšava rad na PHP aplikacijama. Unutar XAMPP-a, koristi se phpMyAdmin za rad s bazama podataka, iako u ovom slučaju baza podataka na kraju nije korištena radi komplikacija sa prebacivanjem na server Sveučilišta Sjever, nego su podaci ispitanika spremljeni u tekstualnu datoteku. Osim toga, na stranici su korištene i skripte povezane s Bootstrapom, popularnim alatom za stilizaciju i funkcionalnost. Bootstrap omogućava izradu responzivnih elemenata poput navigacije, gumba i obrazaca, a u kombinaciji s JavaScriptom pruža funkcionalnosti koje olakšavaju interakciju korisnika s web stranicom.

8.4. Funkcionalnost web stranice

U sklopu istraživanja razvijena je početna stranica "index.php", koja sadrži osnovne informacije o samom istraživanju. Na toj stranici korisnici su mogli pročitati opise istraživanja te klikom na gumb "Osnovne informacije" dobiti objašnjenja o svrsi istraživanja i metodologiji, prikaz prvog ekrana kada se otvori web stranica vidimo na Slici 8.3. Drugi ekran sa osnovnim informacijama prikazan je na Slici 8.4. Pritiskom na gumb "Prijava", korisnici su bili preusmjereni na dio stranice gdje su trebali unijeti godine i spol, to možemo vidjeti na Slici 8.5. Ako polje za unos godina nije bilo ispunjeno, nije bilo moguće nastaviti prema testnom dijelu, odnosno nije se mogao pritisnuti gumb "Testna slika". Kada bi ispitanici unijeli ispravne podatke, otvorila bi se stranica "testna_slika.php", koja je služila kao demonstracija procesa ocjenjivanja. Na stranici "testna_slika.php" korisnicima je prvo prikazana originalna slika, nakon koje je slijedila generirana slika na temelju naredbe "expand" (proširi sliku), to možemo vidjeti na Slici 8.6.



Slika 8.3 Početna stranica istraživanja [izvor autor].



Slika 8.4 Stranica sa osnovnim informacijama [izvor autor].

Vaši podaci

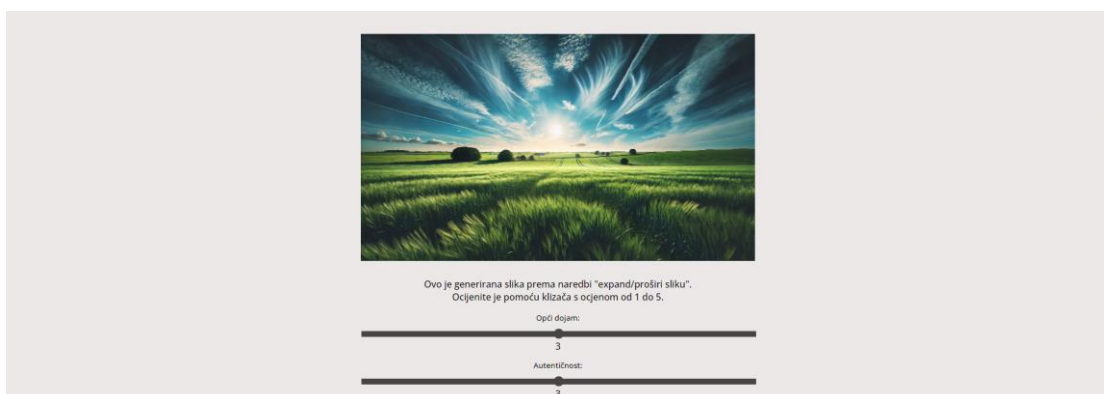
Broj godina:

Spol:

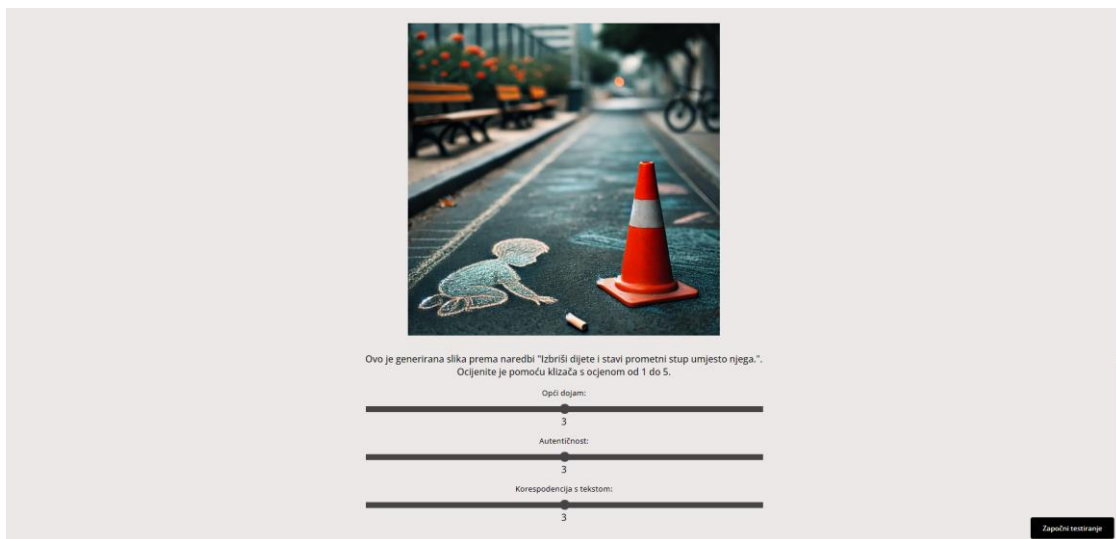
Slika 8.5 Stranica za unos podataka korisnika [izvor autor].



Slika 8.6 Prikaz testne stranice sa originalnom slikom [izvor autor].

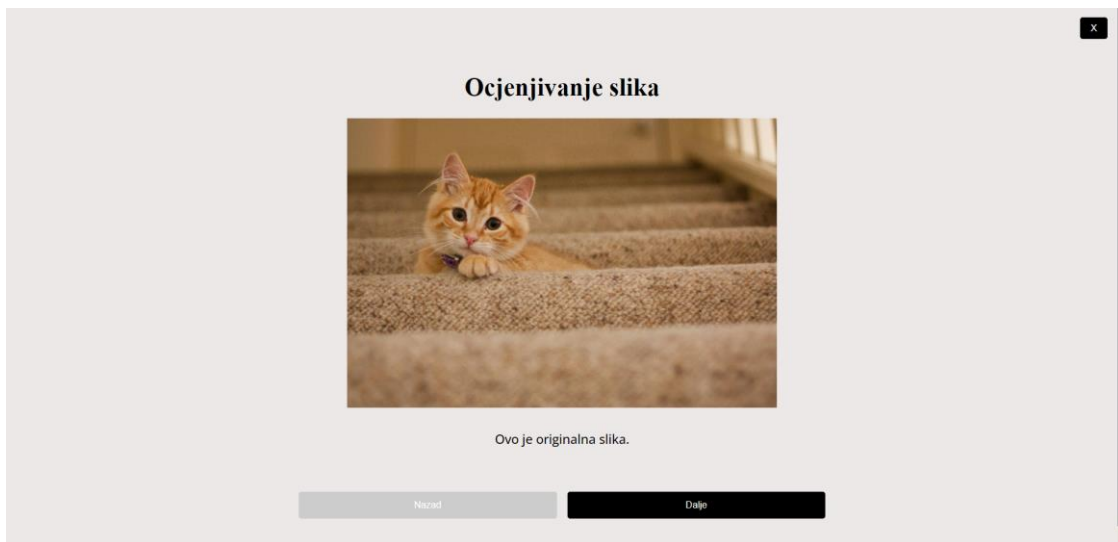


Slika 8.7 Prikaz testne slike prema naredbi „expand/proširi sliku“ [izvor autor].

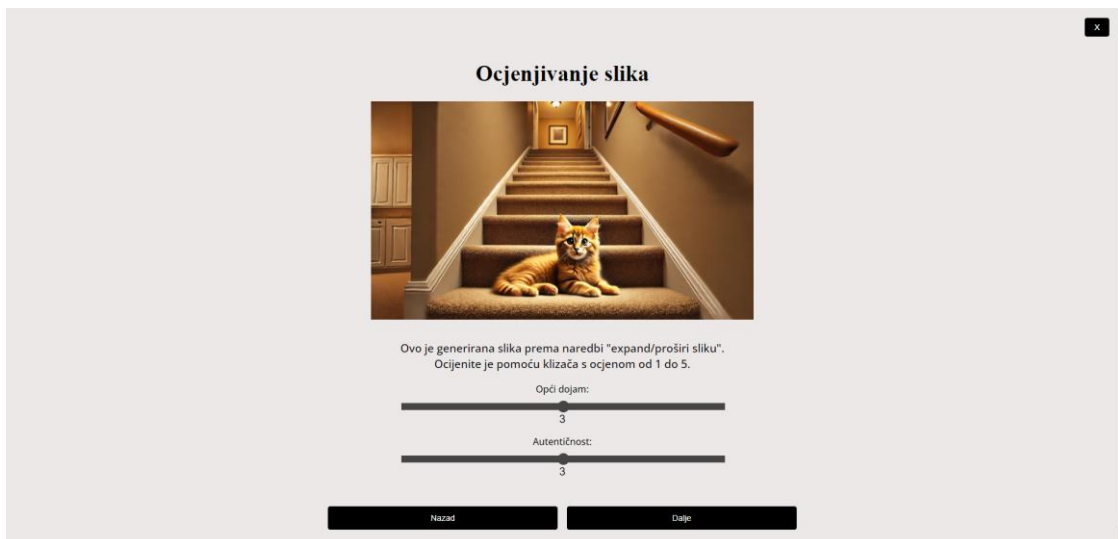


Slika 8.8 Prikaz testne slike prema naredbi „Izbriši dijete i stavi prometni stup umjesto njega“ [izvor autor].

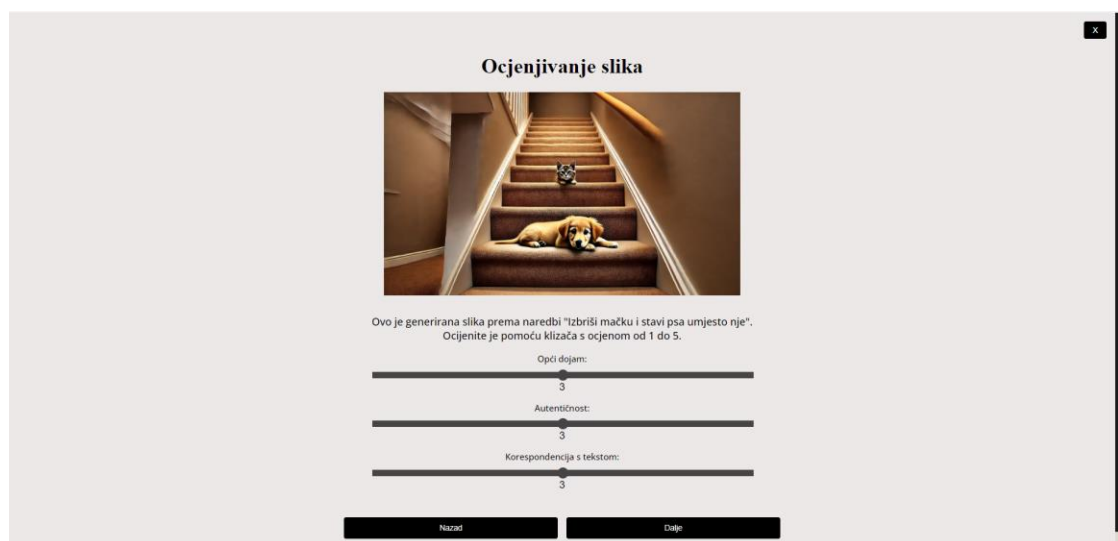
Ispitanici su trebali ocijeniti ovu sliku pomoću klizača s rasponom od 1 do 5, što se može uočiti na Slici 8.7. Nakon toga prikazivala se druga originalna slika, s generiranom verzijom prema naredbi "erase and replace" (izbriši dijete i zamijeni ga prometnim stupom), koju su također trebali ocijeniti na isti način, vidimo to na Slici 8.8. Ova demonstracija omogućila je ispitanicima da steknu uvid u način provođenja ispitivanja te kako koristiti klizače za ocjenjivanje. Navigacija po ovoj stranici bila je omogućena klizačem na desnoj strani ili pomicanjem kotačića miša. Nakon ocjenjivanja testnih slika, pritiskom na gumb "Započni istraživanje" u donjem desnom kutu stranice, ispitanici su prelazili na stranicu "testiranje.php", koja je sadržavala 75 slika. Prvo je bila prikazana originalna slika, nakon čega su slijedile četiri generirane verzije za svaku originalnu sliku. Prikaz originalne slike za prvi primjer vidimo na Slici 8.9. Prve dvije slike generirane su pomoću DALL-E-a, prema naredbama proširi sliku i brisanje i zamjena, što možemo vidjeti na Slici 8.10 i 8.11, a nakon toga su prikazane slike generirane u Midjourney-u. Ispitanici su morali ocijeniti svaku sliku pomoću klizača, prema vizualnoj kvaliteti i usklađenosti s naredbom.



Slika 8.9 Prikaz početnog ekrana na stranici za ocjenjivanje slika [izvor autor].



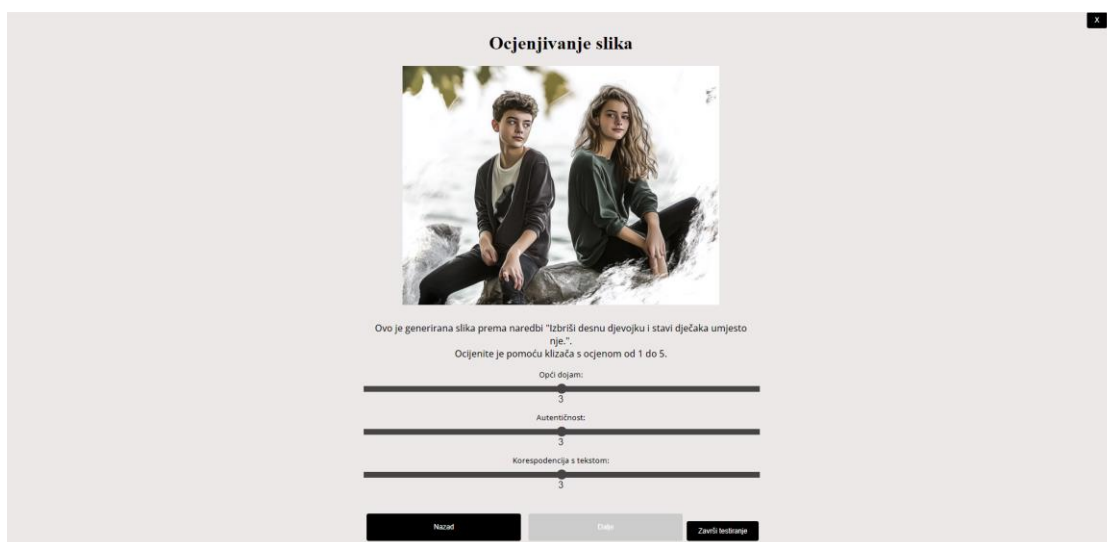
Slika 8.10 Prikaz prve generirane slike prema naredbi „expand/proširi sliku“ sa klizačima za ocjenu [izvor autor].



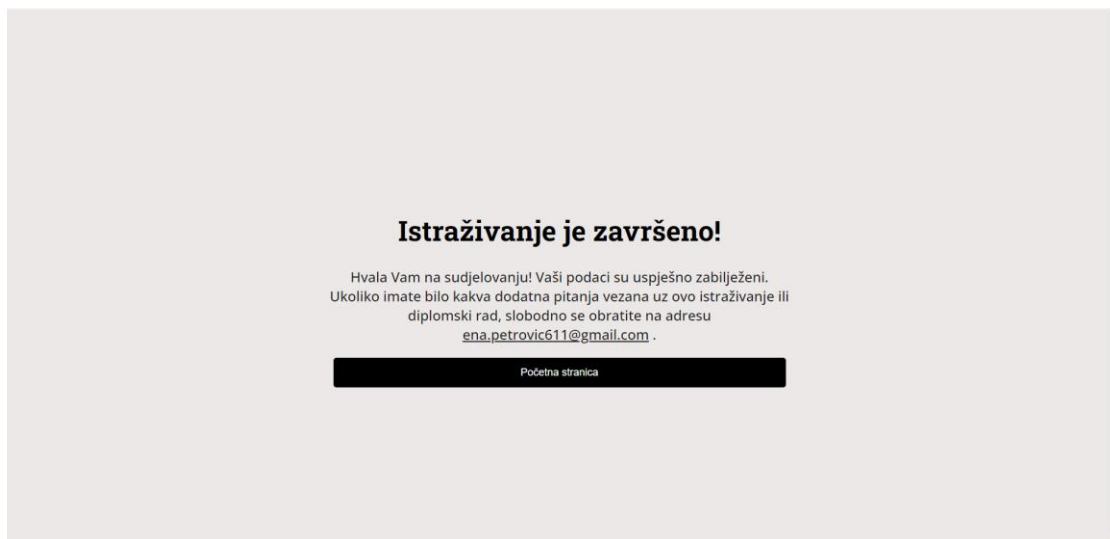
Slika 8.11 Prikaz prve generirane slike prema naredbi „Izbriši mačku i stavi psa umjesto nje“ sa

klizačima za ocjenu [izvor autor].

Navigacija na stranici "testiranje.php" bila je moguća putem klizača na desnoj strani, pomicanjem kotačića miša ili korištenjem gumba "Nazad" i "Dalje" na dnu stranice. Gumb "Nazad" omogućavao je vraćanje na prethodne odgovore, što je bilo korisno ako su ispitanici htjeli promijeniti ocjenu ili ponovno vidjeti originalnu sliku. Gumb "Dalje" služio je za prelazak na sljedeću sliku, pri čemu su se odgovori automatski spremali. Na kraju ispitivanja, nakon ocjenjivanja zadnje slike, pojavio se gumb "Završi istraživanje", koji je vodio na završnu stranicu "zavrсна_stranica.php" i ujedno spremio zadnji odgovor u .txt odgovori_ispitanika. Prikaz zadnje slike istraživanja sa gumbom „Završi istraživanje“ nalazi se na Slici 8.12. Na toj stranici bila je prikazana zahvala ispitanicima na sudjelovanju, pružene su kontakt informacije za daljnja pitanja, te je ponuđen gumb "Početna stranica", koji je ispitanike vraćao na početnu stranicu "index.php". Prikaz zadnje stranice vidimo na Slici 8.13.



Slika 8.12 Prikaz zadnje slike istraživanja te gumba „Završi istraživanje“ [izvor autor].



Slika 8.13 Prikaz zadnje stranice istraživanja [izvor autor].

8.5. Stavljanje web stranice na server

U ovom projektu korišteni su različiti alati za izradu, testiranje i postavljanje istraživačke web stranice. Visual Studio Code, popularan i besplatan uređivač koda, omogućio je pregled cijelog projekta u strukturiranim mapama i olakšao proces pisanja koda zahvaljujući pametnim alatima kao što su automatski prijedlozi koda i zatvaranje zagrada. Za lokalno pokretanje i testiranje PHP jezika koristio se XAMPP, softverski paket koji uključuje Apache web server. Projektne datoteke, uključujući HTML, CSS, JavaScript, PHP i tekstualne datoteke za pohranu podataka, smještene su unutar XAMPP-ove "htdocs" mape. Testiranje je obavljeno putem lokalnog web preglednika, a URL "http://localhost/ moj_projekt/index.php" korišten je za provjeru funkcionalnosti. Za udaljeni pristup računalu na Sveučilištu Sjever korištena je aplikacija AnyDesk, dok je prijenos datoteka na web poslužitelj obavljen putem programa WinSCP, koji omogućuje siguran prijenos datoteka korištenjem SFTP protokola. Prikaz projektnih datoteka u programu WinSCP vidi se na Slici 8.14. Kako bi omogućili spremanje podataka u tekstualnu datoteku (.txt), bilo je potrebno postaviti odgovarajuće dozvole za zapisivanje. To se postiglo omogućavanjem prava pristupa datoteci, čime se osiguralo da sustav može upisivati podatke u datoteku tijekom izvođenja aplikacije. Na kraju, stranica je uspješno postavljena na web poslužitelj Sveučilišta Sjever, dostupna pod URL-om "http://msl.unin.hr/ moj_projekt/index.php".

C:\Users\Student\Desktop\				/var/www/html/moj_projekt/				
Name	Size	Type	Changed	Name	Size	Changed	Rights	Owner
..		Parent directory	9/7/2024 7:44:07 PM	..		9/9/2024 12:59:46 PM	rw-rw-r-x	root
obrada_rezultata		File folder	9/7/2024 7:44:10 PM	Testne slike		9/9/2024 12:50:44 PM	rw-r-xr-x	system1
test_images		File folder	9/5/2024 1:29:43 PM	Originalne slike		9/9/2024 12:50:42 PM	rw-r-xr-x	system1
matlab - Shortcut.lnk	2 KB	Shortcut	9/7/2024 7:43:44 PM	Naredba expand - Mi...		9/9/2024 12:50:38 PM	rw-r-xr-x	system1
				Naredba expand - DA...		9/9/2024 12:50:33 PM	rw-r-xr-x	system1
				Naredba erase and re...		9/9/2024 12:50:27 PM	rw-r-xr-x	system1
				Naredba erase and re...		9/9/2024 12:50:22 PM	rw-r-xr-x	system1
				zavrsna_stanica.php	4 KB	9/3/2024 7:03:24 PM	rw-r--r--	system1
				testna_slika.php	8 KB	9/3/2024 7:30:00 PM	rw-r--r--	system1
				testiranje.php	86 KB	9/4/2024 6:53:06 PM	rw-r--r--	system1
				style.css	3 KB	8/13/2023 4:36:52 PM	rw-r--r--	system1
				skripta.js	12 KB	8/13/2023 7:46:34 PM	rw-r--r--	system1
				odgovori_izpitanka.txt	85 KB	9/6/2024 12:46:34 PM	rw-r--r--	system1
				index.php	10 KB	9/4/2024 7:54:28 PM	rw-r--r--	system1

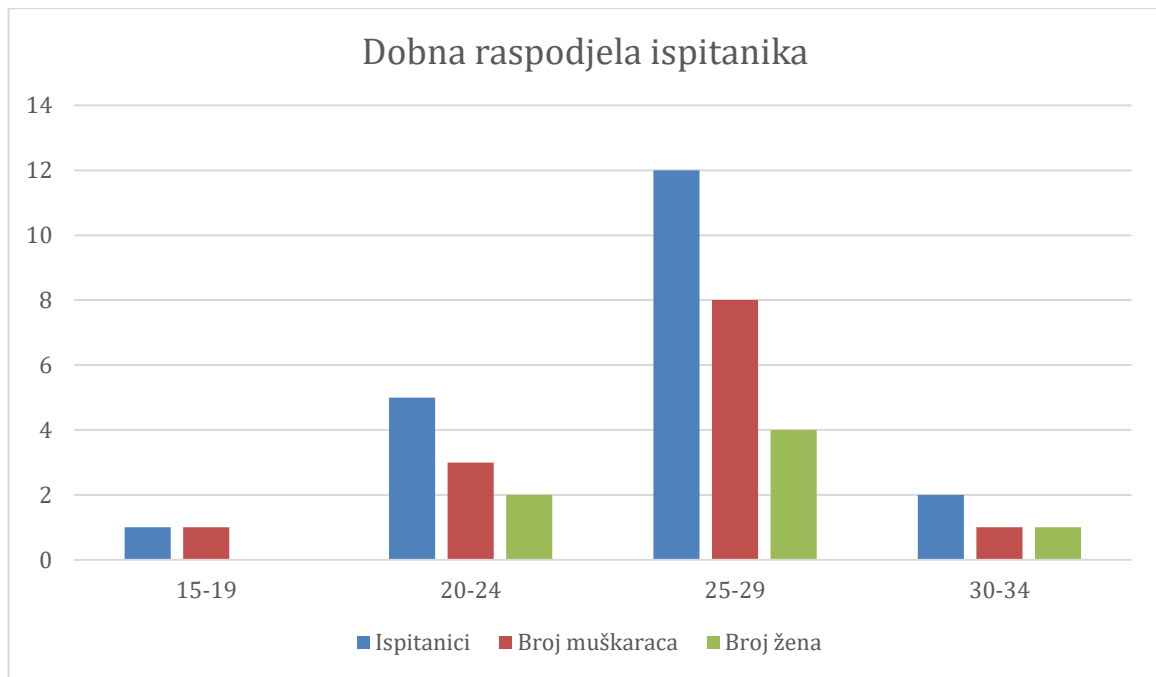
Slika 8.14 Prikaz projektnih datoteka u programu WinSCP [izvor autor].

9. Pregled dobivenih rezultata

Ovo poglavlje započet ćemo s informacijama o subjektivnom istraživanju provedenom na 21 ispitanika, pri čemu je jedna ocjena odbačena zbog nevažćih odgovora. Istraživanje je trajalo pet dana, a detaljnu analizu prikupljenih rezultata predstaviti ćemo u nastavku. Subjektivno ispitivanje provedeno je putem web ankete koju su ispitanici ispunjavali koristeći web preglednik, prema mojim uputama. Stranica s anketom bila je dizajnirana i optimizirana za pregledavanje putem web preglednika, te sam zbog toga zamolila ispitanike da je gledaju isključivo na taj način. Stranica je izrađena u svijetlim tonovima, dok su gumbi bili tamni kako bi se stvorio jasan kontrast i osigurala dobra preglednost. Fotografije korištene za ispitivanje prikazane su u dovoljno velikom formatu kako bi ispitanici mogli što detaljnije uočiti razlike između slika i na taj način dati što preciznije odgovore. Nakon toga, provedeno je objektivno istraživanje pomoću mjera PIQE, TOPIQ_NR i TReS, te je napravljena korelacija između subjektivnih i objektivnih mjera korištenjem Pearsonovog, Spearmanovog i Kendallovog koeficijenta korelacije.

9.1. Pregled osnovnih podataka

Prosječna starost ispitanika iznosila je 25,55 godina. Od ukupno 20 ispitanika, većinu su činili muškarci, njih 13 (65%), dok je 7 ispitanika (35%) bilo ženskog spola. Prikaz dobne raspodjele ispitanika vidi se na Slici 9.1. Prosječno vrijeme potrebno za ispunjavanje upitnika iznosilo je 12 minuta i 39 sekundi. U ovo vrijeme uključeno je i rješavanje testnog pitanja, koje je poslužilo za upoznavanje ispitanika s procesom ocjenjivanja. Važno je napomenuti da se ocjene dobivene na tom testnom pitanju nisu računavale u završnu analizu. Prosječno vrijeme ispunjavanja upitnika kod muškaraca iznosilo je 12 minuta i 29 sekundi, dok je kod žena prosječno trajanje bilo nešto duže, 12 minuta i 58 sekundi.

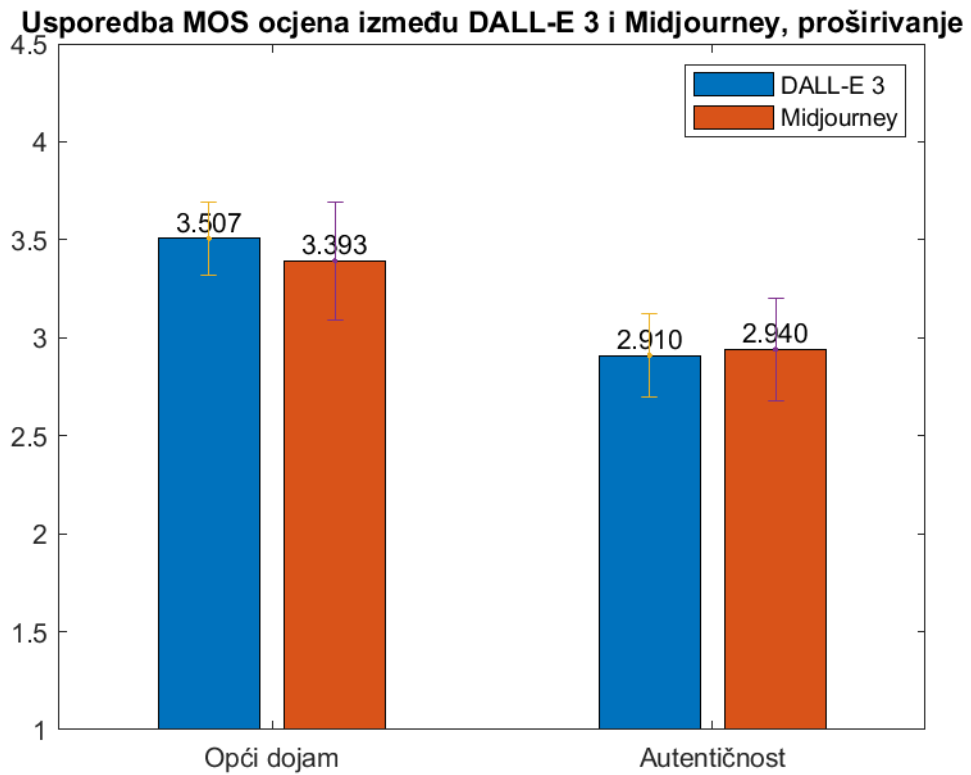


Slika 9.1 Dobna raspodjela ispitanika - graf

9.2. Usporedba subjektivnih ocjena

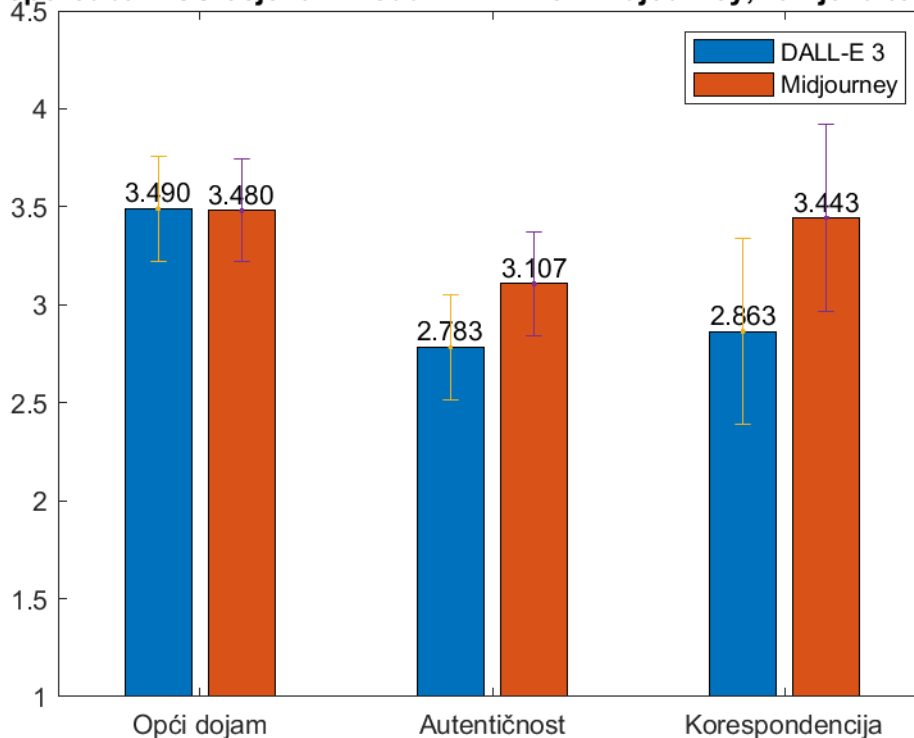
Slika 9.2 prikazuje usporedbu srednje ocjene korisničkog mišljenja (MOS - Mean Opinion Score) između dva alata za generiranje slika, DALL-E 3 i Midjourney. Usporedba se temelji na dvije ključne kategorije: opći dojam i autentičnost. U kategoriji općeg dojma, DALL-E 3 postiže nešto višu ocjenu (3.507) u odnosu na Midjourney (3.393), što sugerira da korisnici imaju blago povoljniji dojam o DALL-E 3 u cjelokupnoj percepciji. Kada je riječ o autentičnosti, rezultati su gotovo izjednačeni. Midjourney postiže blago višu ocjenu (2.940) u usporedbi s DALL-E 3

(2.910), no ta je razlika zanemarljivo mala. Zaključno, iako DALL-E 3 ima prednost u općem dojmu, Midjourney vrlo blago prednjači u autentičnosti, no obje razlike su relativno suptilne.



Slika 9.2 Usporedba MOS ocjena između DALL-E 3 i Midjourney za naredbu proširivanja

Usporedba MOS ocjena između DALL-E 3 i Midjourney, zamjena tekstem

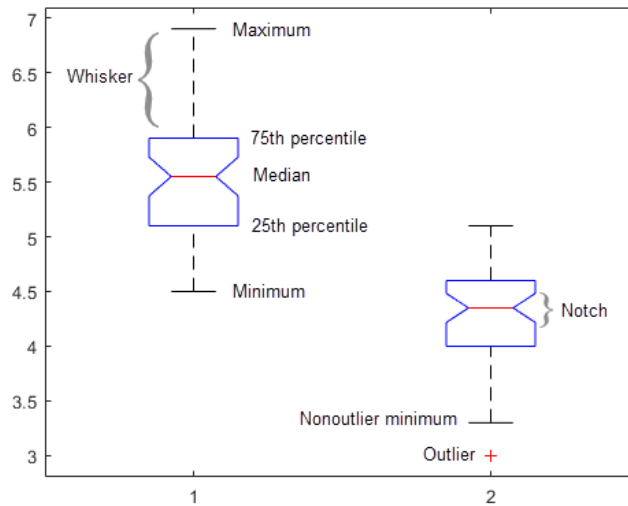


Slika 9.3 Usporedba MOS ocjena između DALL-E 3 i Midjourney, zamjena tekstem

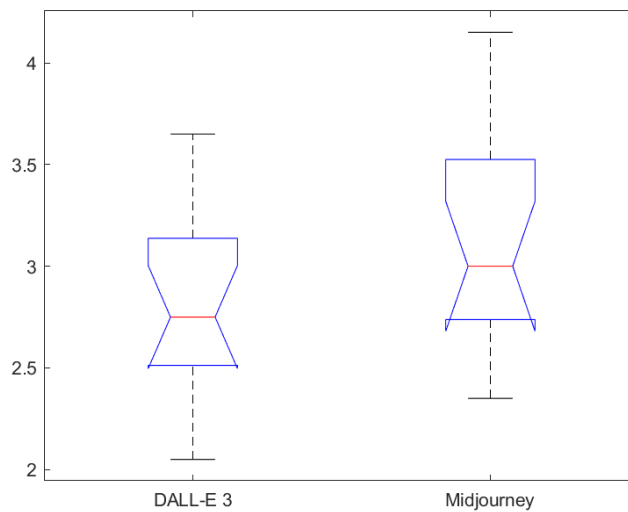
Slika 9.3 prikazuje usporedbu srednje ocjene korisničkog mišljenja (MOS) između alata za generiranje slika DALL-E 3 i Midjourney. S naglaskom na zamjenu teksta na slikama, prikazani su zanimljivi rezultati. Iako su oba alata vrlo slična u pogledu općeg dojma, gdje DALL-E 3 ima ocjenu od 3.490, a Midjourney 3.480, razlike postaju izraženije u drugim aspektima. U kategoriji autentičnosti, Midjourney je nadmašio DALL-E 3, dobivši ocjenu od 3.107 u usporedbi s DALL-E-ovih 2.783. Iako je razlika očita, statistička značajnost ovog rezultata nije potvrđena, budući da je p-vrijednost iznosila 0,0774, što je granično značajno. Sličan trend može se vidjeti i u kategoriji korespondencije, gdje Midjourney ponovno dobiva višu ocjenu 3.443 naspram 2.863 koliko je dobio DALL-E 3. Iako ovaj rezultat ukazuje na bolju preciznost i povezanost između generiranog teksta i slike u Midjourney-u, statistički nije potvrđeno da je razlika značajna (p-vrijednost 0,0719). Ovi rezultati sugeriraju da, iako su oba alata vrlo konkurentna u smislu općeg dojma, Midjourney je prema korisničkim ocjenama bolji u osiguravanju autentičnosti i korespondencije prilikom zamjene teksta na slikama, ali te razlike nisu statistički značajne.

ANOVA test (Analiza varijance) je statistička metoda koja se koristi za ispitivanje razlika između srednjih vrijednosti u dvije ili više grupa podataka. Cilj ANOVA testa je utvrditi je li varijabilnost među grupama veća od varijabilnosti unutar grupa, odnosno može li se zaključiti da su srednje vrijednosti grupa statistički značajno različite. Ako je p-vrijednost iz ANOVA testa manja od unaprijed određene razine značajnosti (obično 0.05), zaključujemo da postoji statistički

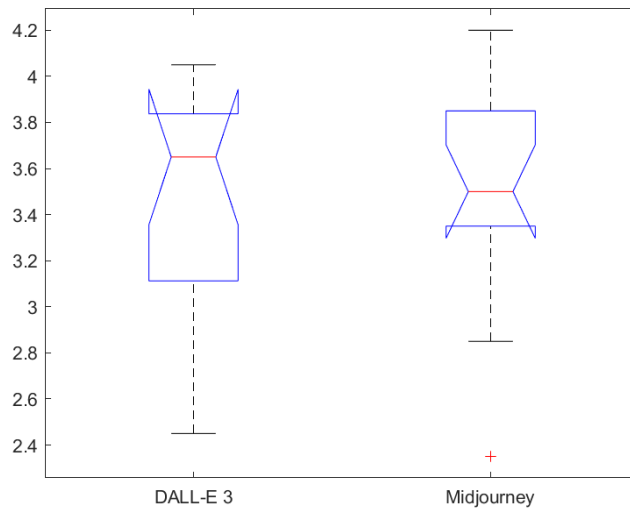
značajna razlika između grupa. U suprotnom, ako je p-vrijednost veća, zaključujemo da nema značajnih razlika između grupa. U kontekstu ANOVA testa, kutijasti dijagram može pomoći vizualizirati razlike između grupa, dok ANOVA test statistički potvrđuje jesu li te razlike značajne. Općeniti izgled kutijastog dijagrama (box plot), koji ilustrira osnovne elemente korištene u analizi podataka nalazi se na Slici 9.4.



Slika 9.4 Općeniti izgled kutastog dijagrama

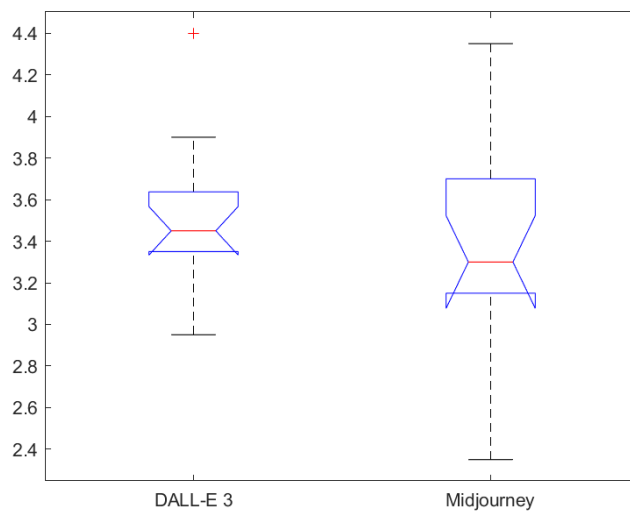


Slika 9.5 Rezultati za autentičnost - zamjena teksta

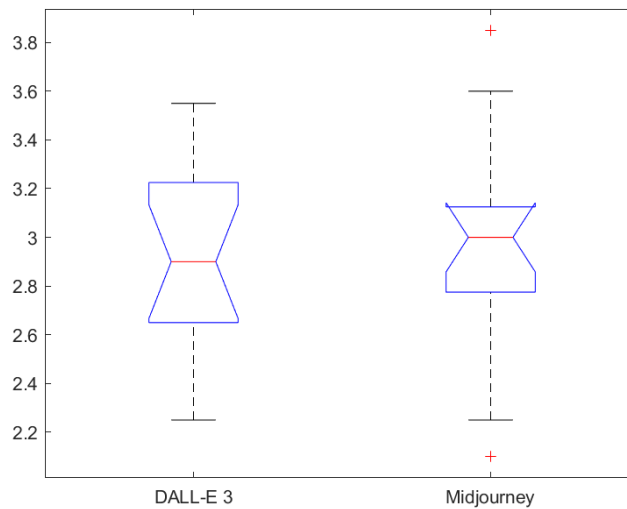


Slika 9.6 Rezultati za opći dojam - zamjena teksta

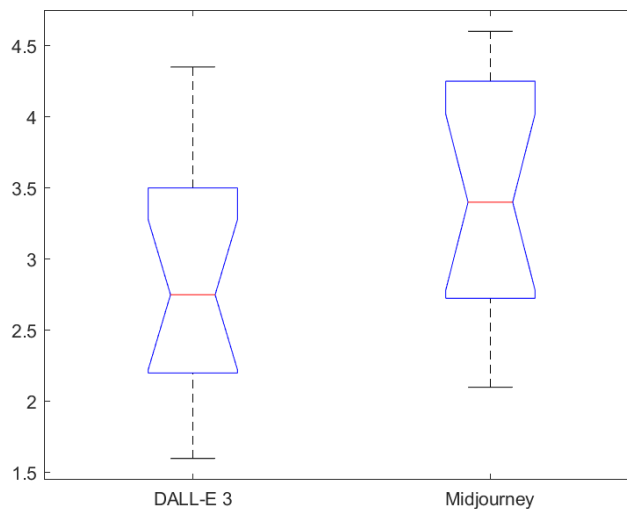
Usporedba prikazana na Slici 9.5 i Slici 9.6 otkriva razlike u ocjenama autentičnosti i općeg dojma između DALL-E 3 i Midjourney alata. Za autentičnost, Midjourney postiže nešto viši medijan ocjena (oko 3.2 naspram 3.0 za DALL-E 3) i pokazuje veću konzistentnost s užim rasponom ocjena. Kod općeg dojma, oba alata imaju sličan medijan (oko 3.6), no Midjourney opet pokazuje manju varijabilnost. Ipak, Midjourney ima jednu iznimno nisku ocjenu, što ukazuje na znatno nižu pojedinačnu ocjenu. DALL-E 3, s druge strane, ima širi raspon ocjena u obje kategorije, što ukazuje na veću raznolikost korisničkih dojmova.



Slika 9.7 Rezultati za opći dojam - proširivanje



Slika 9.8 Rezultati za autentičnost - proširivanje



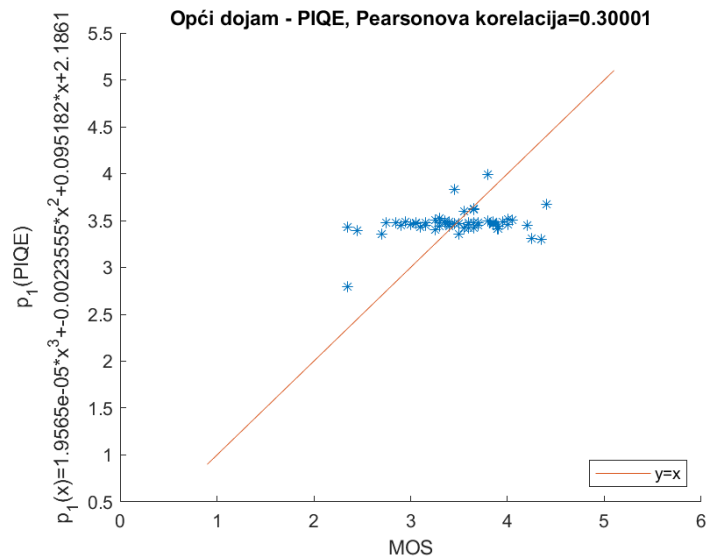
Slika 9.9 Rezultati za korespondenciju - proširivanje

Tri dijagrama prikazana na Slici 9.7, 9.8 i 9.9 uspoređuju rezultate DALL-E 3 i Midjourney alata u kategorijama autentičnosti, općeg dojma i korespondencije. U kategoriji autentičnosti, Midjourney pokazuje nešto viši medijan (oko 3.2 naspram 3.0 za DALL-E 3) i uži raspon ocjena, što ukazuje na konzistentniju percepciju korisnika. Kod općeg dojma, oba alata imaju isti medijan (oko 3.6), ali Midjourney ima veću konzistentnost ocjena, iako postoji jedna značajno niža ocjena. U kategoriji korespondencije, medijani su ponovno slični (oko 3.5), no DALL-E 3 ima jednu izrazito visoku ocjenu, dok Midjourney pokazuje nešto širi raspon ocjena. Općenito, oba alata su vrlo slična, no Midjourney ima nešto konzistentnije ocjene, dok DALL-E 3 pokazuje širi raspon u nekoliko kategorija.

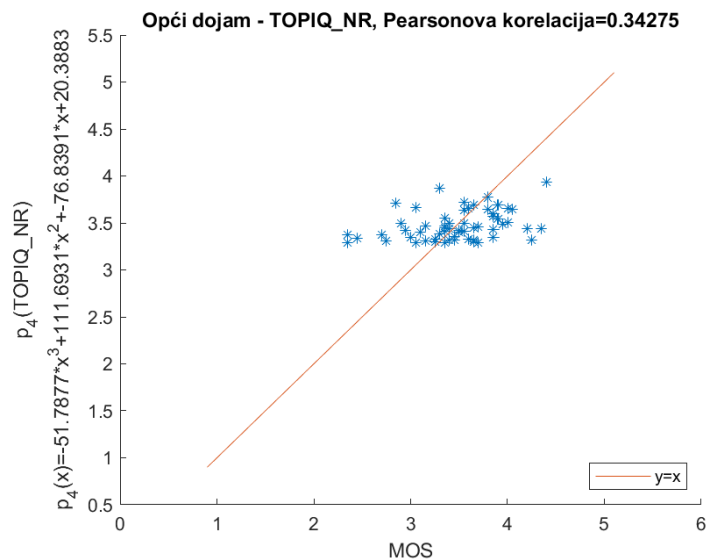
9.3. Korelacija subjektivnih i objektivnih mjera

Korelacija između subjektivnih i objektivnih mjera odnosi se na povezanost između osobnih dojmova korisnika i mjerljivih podataka. Subjektivne mjere temelje se na individualnim ocjenama, dok se objektivne mjere odnose na kvantitativne, mjerljive podatke, poput tehničke izvedbe alata. Kada postoji pozitivna korelacija, poboljšanja u objektivnim mjerama prate bolje subjektivne ocjene. Negativna korelacija znači da se subjektivne ocjene smanjuju s rastom objektivnih mjera, dok niska korelacija ukazuje na malu povezanost između ova dva tipa mjera. U ovom istraživanju korelacija između objektivnih i subjektivnih mjera je niska za sve mjere, te ćemo sada u nastavku detaljnije pojasniti rezultate.

Rezultati Pearsonove korelacije pokazuju stupanj povezanosti između objektivnih mjera (PIQE, TOPIQ_NR, TReS) i subjektivnih ocjena (Opći dojam, Autentičnost, Korespondencija). U kategoriji općeg dojma, korelacije su blage do umjerene, s TOPIQ_NR-om koji pokazuje najjaču povezanost (0,34), dok PIQE i TReS imaju nešto slabije korelacije (0,30 i 0,22). To znači da objektivna poboljšanja u ovim mjerama donekle prate bolji subjektivni dojam korisnika. U kategoriji autentičnosti, korelacije su općenito niske. PIQE (0,18), TOPIQ_NR (0,15) i TReS (0,10) pokazuju vrlo slabu povezanost s percepcijom autentičnosti, što sugerira da objektivna mjerenja malo utječu na subjektivne ocjene autentičnosti. S druge strane, u kategoriji korespondencije, korelacije su najjače, posebno kod PIQE-a (0,43) i TReS-a (0,39), što ukazuje na veću povezanost između ovih mjera i percepcije korisnika. TOPIQ_NR ima znatno slabiju korelaciju (0,18), što znači da promjene u ovoj mjeri imaju manji utjecaj na subjektivne ocjene korespondencije. Dok su korelacije u općem dojmu, što možemo vidjeti na Slici 9.10 i 9.11; i autentičnosti relativno slabe, kod korespondencije je vidljiva značajnija povezanost između objektivnih i subjektivnih mjera, osobito za PIQE i TReS.



Slika 9.10 Opći dojam - PIQE, Pearsonova korelacija



Slika 9.11 Opći dojam - TOPIQ_NR, Pearsonova korelacija

Rezultati Spearmanove korelacije: u kategoriji općeg dojma, korelacija s TOPIQ_NR-om je umjereno pozitivna (0,33), dok su PIQE (0,16) i TReS (0,17) povezani slabije. To znači da poboljšanja u objektivnim mjerama, posebno TOPIQ_NR, donekle utječu na subjektivne ocjene općeg dojma. Za autentičnost, sve korelacije su vrlo slabe. PIQE (0,15), TOPIQ_NR (0,08) i TReS (0,07) pokazuju minimalan utjecaj na subjektivne ocjene autentičnosti, što sugerira da ove objektivne mjere imaju mali utjecaj na korisničku percepciju autentičnosti. Kategorija korespondencije pokazuje zanimljive rezultate. Korelacija s PIQE-om je umjereno negativna (-0,36), što znači da poboljšanja u PIQE-u prate smanjenje subjektivnih ocjena korespondencije. Korelacije s TOPIQ_NR (0,04) i TReS (-0,04) su vrlo slabe, što ukazuje na gotovo zanemarivu povezanost s ocjenama korespondencije. Spearmanova korelacija pokazuje blage do umjerene

povezanosti između objektivnih mjera i subjektivnih ocjena, s najjačom pozitivnom korelacijom u općem dojmu, dok je u korespondenciji zabilježena negativna povezanost s PIQE-om.

Rezultati Kendallove korelacije: u kategoriji općeg dojma, TOPIQ_NR pokazuje umjerenu pozitivnu korelaciju (0,23), što ukazuje na značajniju povezanost između ove mjere i subjektivnih ocjena. PIQE (0,11) i TReS (0,12) imaju slabije, ali pozitivne korelacije, što sugerira blagu povezanost s općim dojmom. Za autentičnost, sve korelacije su vrlo slabe, pri čemu PIQE (0,11), TOPIQ_NR (0,06) i TReS (0,06) imaju zanemariv utjecaj na subjektivne ocjene autentičnosti, što ukazuje na mali utjecaj ovih mjerenja na percepciju autentičnosti. Kategorija korespondencije pokazuje zanimljiv rezultat, s umjereno negativnom korelacijom između PIQE-a i subjektivnih ocjena (-0,28). To znači da kako se PIQE poboljšava, ocjene korespondencije opadaju. S druge strane, korelacije s TOPIQ_NR (0,04) i TReS (-0,04) su vrlo slabe, što sugerira da ove mjere imaju minimalan utjecaj na subjektivne ocjene korespondencije. Kendallova korelacija pokazuje da objektivne mjere imaju slabu do umjerenu povezanost sa subjektivnim ocjenama, s najjačom korelacijom u kategoriji općeg dojma, dok je povezanost u kategorijama autentičnosti i korespondencije znatno slabija, osim umjereno negativne korelacije s PIQE-om u korespondenciji.

Iz rezultata se može zaključiti da su korelacije između subjektivnih i objektivnih mjera općenito slabe, ali se razlikuju ovisno o kategoriji. Opći dojam pokazuje najjaču pozitivnu korelaciju, posebno s TOPIQ_NR-om, što znači da promjene u ovoj objektivnoj mjeri donekle prate promjene u subjektivnim ocjenama. U kategoriji autentičnosti, sve korelacije su vrlo slabe, sugerirajući da objektivne mjere nemaju veliki utjecaj na korisničku percepciju autentičnosti. Kategorija korespondencije pokazuje zanimljiv fenomen - PIQE je negativno povezan sa subjektivnim ocjenama, što znači da poboljšanja u PIQE-u prate smanjenje ocjena korespondencije. Ovaj obrazac se javlja u Spearmanovoj i Kendallovoj korelaciji, dok su TOPIQ_NR i TReS vrlo slabo povezani s korespondencijom. Općenito, analiza pokazuje da objektivne mjere samo djelomično prate subjektivne ocjene korisnika, pri čemu je povezanost u nekim slučajevima slabija ili čak negativna, osobito kod PIQE-a u kategoriji korespondencije.

10. Zaključak

Budućnost umjetne inteligencije u generiranju slika i grafičkom dizajnu obećava značajne promjene. Tehnologije poput generativnih modela, dubokog učenja i difuzijskih mreža brzo napreduju, omogućujući stvaranje vizualnog sadržaja s visokim stupnjem realističnosti i kreativnosti. Istraživanje ovog rada je analiziralo kvalitetu slika koje su nadopunjene pomoću modela dubokog učenja, i to kombiniranjem objektivnih mjera i subjektivnih procjena korisnika. Cilj je bio ispitati koliko se rezultati objektivnih mjera slažu s onim što korisnici stvarno doživljavaju kad gledaju slike. U kategorijama autentičnosti i usklađenosti slike s tekstualnim opisom, te povezanosti su bile vrlo slabe, pa čak i negativne. Rezultati pokazuju da objektivne mjere, iako korisne, ne mogu uvijek prikazati kako ljudi percipiraju kvalitetu slika. Ljudi ocjenjuju slike na temelju emocija, konteksta i vlastitih preferencija, što objektivne mjere ne uzimaju u obzir. Zbog toga bi bilo korisno razviti nove metode koje bi bolje pratile ljudsku percepciju. Zaključno, ovaj rad ističe važnost uzimanja u obzir i objektivnih i subjektivnih faktora prilikom razvoja modela za nadopunjavanje slika. To bi moglo pomoći u stvaranju slika koje ne samo da izgledaju bolje, već su i korisnicima vizualno ugodnije i smislenije.

11. Literatura

- [1] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review", *Computational Intelligence and Neuroscience*, 2018, 7068349, str. 1-13, 2018. doi: 10.1155/2018/7068349
- [2] R. Szeliski, "Computer Vision: Algorithms and Applications", 2nd ed., Springer Cham, 2023., doi: 10.1007/978-3-030-34372-9
- [3] R. Gonzalez i R. Woods, "Digital Image Processing", 4th ed., Prentice Hall, Upper Saddle River, N.J., 2017.
- [4] W. K. Mutlag, S. K. Ali, Z. M. Aydam i B. H. Taher, "Feature Extraction Methods: A Review", *Journal of Physics Conference Series*, vol. 1591, 012028, 2020., doi:10.1088/1742-6596/1591/1/012028
- [5] B. Mirzaei, H. Nezamabadi-pour, A. Raoof i R. Derakhshani, "Small Object Detection and Tracking: A Comprehensive Review", *Sensors*, vol. 23, no. 15: 6887, 2023., doi: 10.3390/s23156887
- [6] J. Salvi, X. Armangué i J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation", *Pattern Recognition*, vol 35, no. 7, 2002., doi: 10.1016/S0031-3203(01)00126-1
- [7] F. Rameau, J. Park, O. Bailo, I. S. Kweon i MC-Calib, "A generic and robust calibration toolbox for multi-camera systems", *Computer Vision and Image Understanding*, vol. 217, 103353, 2022., doi: 10.1016/j.cviu.2021.103353
- [8] N. Sharmin i R. Brad., "Optimal Filter Estimation for Lucas-Kanade Optical Flow", *Sensors*, vol. 2012, no. 9: 12694-12709., 2012., doi: 10.3390/s120912694
- [9] J. Ho i T. Salimans, "Classifier-free diffusion guidance "
- [10] Prafulla Dhariwal, Alex Nichol - Diffusion Models Beat GANs on Image Synthesis - arXiv:2105.05233
- [11] S. Xie, Z. Zhang, Z. Lin, T. Hinz i K. Zhang, "SmartBrush: Text and Shape Guided Object Inpainting with Diffusion Model," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Kanada, 2023., str. 22428-22437, doi: 10.1109/CVPR52729.2023.02148.
- [12] Y. Zeng, Z. Lin, J. Zhang, Q. Liu, J. Collomosse, J. Kuen, V. M. Patel, "SceneComposer: Any-Level Semantic Image Synthesis," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Kanada, 2023., str. 22468-22478, doi: 10.1109/CVPR52729.2023.02152.
- [13] T. Anciukevičius, et al., "Renderdiffusion: Image diffusion for 3D reconstruction, inpainting and generation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
- [14] S. Wang et al., "Imagen Editor and EditBench: Advancing and Evaluating Text-Guided Image Inpainting", 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Kanada, str. 18359-18369, 2023., doi: 10.1109/CVPR52729.2023.01761
- [15] A. B. Yildirim, V. Baday, E. Erdem, A. Erdem, A. Dundar, "Inst-Inpaint: Instructing to Remove Objects with Diffusion Models", arXiv:2304.03246v2, 2023., doi: 10.48550/arXiv.2304.03246

- [16] Y. -C. Cheng, C. H. Lin, H. -Y. Lee, J. Ren, S. Tulyakov i M. -H. Yang, "InOut: Diverse Image Outpainting via GAN Inversion," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, SAD, str. 11421-11430, 2022., doi: 10.1109/CVPR52688.2022.01114
- [17] H. Shi et al., "FishDreamer: Towards Fisheye Semantic Completion via Unified Image Outpainting and Segmentation," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Kanada, str. 6434-6444, 2023., doi: 10.1109/CVPRW59228.2023.00684
- [18] P. Mohammadi, A. Ebrahimi-Moghadam, S. Shirani, "Subjective and Objective Quality Assessment of Image: A Survey", arXiv:1406.7799, 2014., doi: 10.48550/arXiv.1406.7799
- [19] ITU-R BT.500-15., "BT.500 : Methodologies for the subjective assessment of the quality of television images", International Telecommunications Union, 2023.
- [20] ITU-T Technical Report, " PSTR-CROWDS Subjective evaluation of media quality using a crowdsourcing approach ", International Telecommunications Union, 2018.
- [21] ITU-T Focus Group on Application of Distributed Ledger Technology (FG DLT), "Technical Report FG DLT D2.1 Distributed ledger technology use cases ", International Telecommunications Union, 2019.
- [22] C. Li et al., "AGIQA-3K: An Open Database for AI-Generated Image Quality Assessment," IEEE Transactions on Circuits and Systems for Video Technology, 2023., doi: 10.1109/TCSVT.2023.3319020.
- [23] S. A. Golestaneh, S. Dadsetan i K. M. Kitani, "No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency", 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, SAD, str. 3989-3999, 2022., doi: 10.1109/WACV51458.2022.00404
- [24] Y. K. Yun i W. Lin, "You Only Train Once: A Unified Framework for Both Full-Reference and No-Reference Image Quality Assessment ", arXiv:2310.09560v2, doi: 10.48550/arXiv.2310.09560
- [25] Su et al., "Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, SAD, str. 3664-3673, 2020., doi: 10.1109/CVPR42600.2020.00372
- [26] N. Venkatanath, D. Praneeth, Bh. Maruthi Chandrasekhar, S. S. Channappayya i S. S. Medasani, "Blind image quality evaluation using perception based features", 2015 Twenty First National Conference on Communications (NCC), Mumbai, Indija, str. 1-6, 2015., doi: 10.1109/NCC.2015.7084843
- [27] A. Mittal, R. Soundararajan i A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," IEEE Signal Processing Letters, vol. 20, no. 3, str. 209-212, 2013., doi: 10.1109/LSP.2012.2227726
- [28] C. Chen et al., "TOPIQ: A Top-Down Approach From Semantics to Distortions for Image Quality Assessment," IEEE Transactions on Image Processing, vol. 33, 2024., str. 2404-2418, doi: 10.1109/TIP.2024.3378466
- [29] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, " Deep Unsupervised Learning using Nonequilibrium Thermodynamics ", Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), Lille, Francuska, str.2256-2265, 2015.
- [30] Jonathan Ho, Ajay Jain, Pieter Abbeel, " Denoising Diffusion Probabilistic Models ", 34th Conference on Neural Information Processing Systems (NeurIPS 2020), online, str. 6840-6851, 2020.

- [31] Y. Song i S. Ermon, "Generative modeling by estimating gradients of the data distribution", 33th Conference on Neural Information Processing Systems (NeurIPS 2019), Red Hook, NY, SAD, str. 11918–11930., 2019.
- [32] "Umjetna inteligencija," Hrvatska enciklopedija, mrežno izdanje, Leksikografski zavod Miroslav Krleža, 2013.–2024., dostupno na: <https://www.enciklopedija.hr/clanak/umjetna-inteligencija>
- [33] K. I. Roumeliotis i N. D. Tselikas, "ChatGPT and Open-AI Models: A Preliminary Review," *Future Internet*, vol. 15, no. 6, str. 192, 2023., doi: 10.3390/fi15060192
- [34] G. Marcus, E. Davis i S. Aaronson, "A Very Preliminary Analysis of DALL-E 2," arXiv preprint arXiv:2204.13807, 2022., arXiv:2204.13807.
- [35] "DALL·E 2," OpenAI, dostupno na: <https://openai.com/index/dall-e-2/>
- [36] "Midjourney," Wikipedia, dostupno na: <https://en.wikipedia.org/wiki/Midjourney>
- [37] M. D. Murray, "Generative AI Art: Copyright Infringement and Fair Use," *SMU Sci. & Tech. L. Rev.*, vol. 26, str. 259, 2023.
- [38] "Midjourney," dostupno na: <https://www.midjourney.com/imagine>
- [39] "Silhouette of Man Standing on Mountain Cliff," Pexels, dostupno na: <https://www.pexels.com/photo/silhouette-of-man-standing-on-mountain-cliff-1087735/>

Popis slika

Slika 2.1 Reprezentativni primjerci malih objekata [5].	20
Slika 2.2 Usporedba inverznog i Wienerovog filtriranja. (a) Rezultat inverznog filtriranja slike na slici, (b) Rezultat inverznog filtriranja s radijalnim ograničenjem, (c) Rezultat korištenjem Wienerovog filtra [3].	21
Slika 2.3 CNN arhitektura za klasifikaciju slika [1].	25
Slika 3.1 Generiranje rezultata visokokvalitetnog nadopunjavanja objekata [12].	27
Slika 3.2 korisnike je upitano da odaberu generiranu sliku koja najbolje odgovara maski i unesenom tekstualnom opisu te izgleda najrealističnije [12].	28
Slika 3.3 Primjeri kontrole preciznosti s promptom „astronaut“ [12].	29
Slika 3.4 Rezultati s različitim razinama preciznosti. Za svaki ulazni raspored, uzorkovane su slike počevši od istog šuma, tako da slike na različitim razinama preciznosti mogu imati slične stilove [13].	31
Slika 3.5 Nadopunjavanje unutar slike svjesno 3D prostora [14].	32
Slika 3.6 Niz uređivanja pomoću programa Imagen Editor [15].	33
Slika 3.7 Nadopunjavanje slika instrukcijama [16].	34
Slika 4.1 Nadopunjavanje izvan slike na više različitih primjera, predložena metoda može tražiti različita rješenja za određeni ulaz, postižući visoku raznolikost rezultata bez žrtvovanja kvalitete generiranja [17].	35
Slika 4.2 Ilustracija Fisheye Semantic Completiona [17].	37
Slika 4.3 (gore-lijeva) - obzirom na ulaznu sliku i istrenirani generator, predloženi algoritam traži latentne kodove koji mogu generirati slike koje sadrže ulaznu sliku. Na taj način možemo prirodno postići (gore-desno) raznoliko nadopunjavanje slike, (sredina-desno) kategorijsku manipulaciju za područje nadopunjavanja, i (dolje) generirati panoramu s bogatom i složenom strukturom [19].	38
Slika 5.1 Preporuka ITU-R BT.500 struktura [19].	42
Slika 5.2 Opći tijek rada platformi za mikro zadatke masovnog ispitivanja [20].	43
Slika 5.3 Uzorci slika iz AGIQA-3K baze podataka, gdje prvi do šesti redak prikazuju slike generirane pomoću modela (AttnGAN, DALLÉ2, GLIDE, Midjourney, Stable Diffusion i Stable Diffusion XL), dok stupci označavaju isti ulazni prompt [22].	45
Slika 5.4 (engl. artificial general intelligence, AGI), gdje su ključne riječi za generiranje označene u gornjem desnom kutu. (a) Tehnički problemi su niskorazinska izobličenja poput zamućenja i besmislenih objekata. (b) Artefakti umjetne inteligencije znače neočekivane artefakte kao što su nedostajući/višak udova. (c) Deepfake se odnosi na AGI slike koje HVS prepoznaje	

kao lažne zbog nerealnih proporcija, osvjetljenja, itd. (d) Estetski aspekti ukazuju na estetsku kvalitetu uključujući nedostatak detalja, neugodne objekte itd. [22].	45
Slika 6.1 Okvir RR-IQA sustava za procjenu kvalitete slike s reduciranom referencom [18].	46
Slika 6.2 Ilustracija osjetljivosti NR-IQA modela na horizontalno okretanje. Na desnoj strani svake slike, prikazana je subjektivna ocjena kvalitete (MOS/DMOS) i predviđena ocjena kvalitete; crveni brojevi u zagradama pokazuju apsolutnu razliku između predikcija kada je slika okrenuta [23].	48
Slika 6.3 JP2K izobličena slika iz LIVE baze podataka s dodatnim AWGN šumom i PIQE prostornom maskom kvalitete. Crveno, žuto i zeleno ispunjeni blokovi označavaju blokove koji zadovoljavaju NDC, NC i uniforme kriterije [27].	50
Slika 6.4 Proces proračunavanja predložene mjere HyperIQA [25].	50
Slika 6.5 Ilustracija glavne svrhe predložene mreže (donji dio slike), gdje se različiti parovi slika mogu unositi u mrežu kako bi se dobili FR/NR IQA rezultati koristeći istu arhitekturu. Tradicionalni način izračunavanja kvalitete slike prikazan je kao gornji dio slike radi jednostavnije usporedbe [24].	51
Slika 6.6 Vizualizacija pažnje s različitim vrstama izobličenja iz TID2013 skupa podataka. Prvi redak: Maske modula ograničenog lokalnog udruživanja (engl. gated local pooling, GLP) na 4 razine; Drugi redak: CSA težine [28].	53
Slika 7.1 Prikaz funkcionalnosti alata DALL-E 2 [35].	56
Slika 7.2 Prikaz funkcionalnosti alata Midjourney [38].	57
Slika 8.1 Sučelje web stranice Pexels.com [39].	60
Slika 8.2 Originalna fotografija korištena za generiranje naredbe prikazane u Tablici 8.1 [39].	61
Slika 8.3 Početna stranica istraživanja [izvor autor].	63
Slika 8.4 Stranica sa osnovnim informacijama [izvor autor].	63
Slika 8.5 Stranica za unos podataka korisnika [izvor autor].	64
Slika 8.6 Prikaz testne stranice sa originalnom slikom [izvor autor].	64
Slika 8.7 Prikaz testne slike prema naredbi „expand/proširi sliku“ [izvor autor].	64
Slika 8.8 Prikaz testne slike prema naredbi „Izbriši dijete i stavi prometni stup umjesto njega“ [izvor autor].	65
Slika 8.9 Prikaz početnog ekrana na stranici za ocjenjivanje slika [izvor autor].	66
Slika 8.10 Prikaz prve generirane slike prema naredbi „expand/proširi sliku“ sa klizačima za ocjenu [izvor autor].	66
Slika 8.11 Prikaz prve generirane slike prema naredbi „Izbriši mačku i stavi psa umjesto nje“ sa klizačima za ocjenu [izvor autor].	66
Slika 8.12 Prikaz zadnje slike istraživanja te gumba „Završi istraživanje“ [izvor autor].	67

Slika 8.13 Prikaz zadnje stranice istraživanja [izvor autor].	68
Slika 8.14 Prikaz projektnih datoteka u programu WinSCP [izvor autor].	69
Slika 9.1 Dobna raspodjela ispitanika - graf.....	71
Slika 9.2 Usporedba MOS ocjena između DALL-E 3 i Midjourney za naredbu proširivanja.....	72
Slika 9.3 Usporedba MOS ocjena između DALL-E 3 i Midjourney, zamjena tekstem	73
Slika 9.4 Općeniti izgled kutastog dijagrama	74
Slika 9.5 Rezultati za autentičnost - zamjena teksta.....	74
Slika 9.6 Rezultati za opći dojam - zamjena teksta	75
Slika 9.7 Rezultati za opći dojam - proširivanje.....	75
Slika 9.8 Rezultati za autentičnost - proširivanje	76
Slika 9.9 Rezultati za korespodenciju - proširivanje	76
Slika 9.10 Opći dojam - PIQE, Pearsonova korelacija.....	78
Slika 9.11 Opći dojam - TOPIQ_NR, Pearsonova korelacija	78

Popis tablica

Tablica 8.1 Prikaz rezultata naredbe "brisanje i zamjena" u DALL-E i Midjourney alatima [izvor autor].....	62
--	----