

Bojenje crno-bijelih slika

1st Luka Mucko 2nd Filip Pankretić 3rd Dominik Jambrović 4th Velimir Kovačić 5th Filip Perković 6th Luka Glavinić
FER FER FER FER FER FER

I. UVOD

Bojenje crno-bijelih slika zadatak je u području računalnog vida, a glavna mu je primjena restauracija povijesnih fotografija kako bi njihov prikaz bio što vjerniji stvarnosti. Tradicionalne metode bojenja slika iziskuju značajnu količinu ručnog rada, stručnosti i vremena. S razvojem dubokog učenja i konvolucijskih neuronskih mreža, postalo je moguće automatizirano bojenje slika, što smanjuje potrebu za ručnim radom i omogućuje bržu i učinkovitiju obradu slika. U ovom radu bit će opisano korištenje generativnih modela, posebno generativnih suparničkih mreži (engl. *Generative Adversarial Networks* - GAN), za bojanje crno-bijelih slika. Ta metoda uključuje dva modela: generator, koji stvara obojene slike iz nasumičnog vektora, i diskriminator, koji pokušava razlikovati generirane slike od stvarnih. Specifično je opisano korištenje uvjetnog GAN-a (engl. *conditional GAN* - cGAN), koji na ulazu diskriminatora i generatora dodaje crno-bijelu reprezentaciju slike kao uvjetni vektor. Ovakav pristup omogućuje generiranje visoko kvalitetnih obojenih slika koje su u većini slučajeva vrlo slične onim stvarnim.

II. PREGLED POSTOJEĆIH PRISTUPA

Razne su tehnike i metode bojenja crno-bijelih slika nastajale tijekom godina. Tradicionalne metode bojenja oslanjale su se na ručni unos i različite algoritme za propagaciju boje. Jedan je primjer *Scribble-Based Colorization* [1], gdje korisnik dodaje boje na određene dijelove slike, a algoritam zatim te boje širi po slici.

S razvojem dubokog učenja, došlo je do značajnog napretka u automatizaciji procesa bojenja. Rana istraživanja koristila su konvolucijske neuronske mreže (CNN), pri čemu je mreža učila mapirati crno-bijele slike u obojene slike na temelju velikog skupa obojenih slika. Jedan primjer te metode predstavlja istraživanje *Colorful Image Colorization* [2] iz 2016., u kojem su autori koristili konvolucijsku neuronsku mrežu za predikciju distribucije boja za svaki piksel crno-bijele slike. Pojavom GAN-ova i cGAN-ova kvaliteta generiranih obojenih slika ponovno je znatno poboljšana. Rad *Image-to-Image Translation with Conditional Adversarial Networks* [3] iz 2017. ilustrira korištenje cGAN-a za različite zadatke translacije sa slike na sliku, uključujući i bojanje crno-bijelih slika.

Nedavno uvedene nadogradnje uključuju korištenje *U-Net* arhitekture neuronske mreže u generatoru, koja omogućuje bolje dohvaćanje konteksta i detalja u slici dodavanjem prekočnih veza (engl. *skip connections*) između slojeva kodera

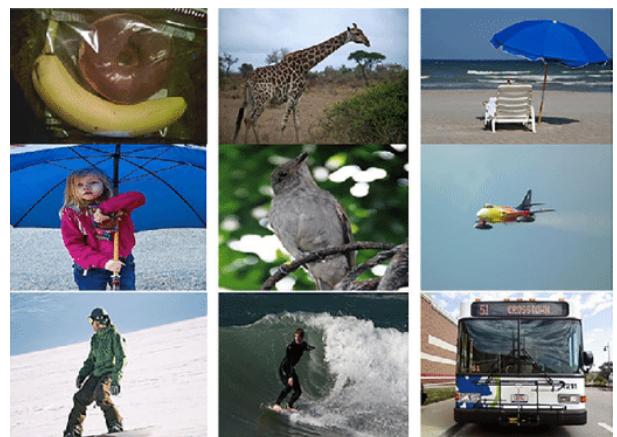
i dekodera. To dovodi do kvalitetnijih rezultata, posebno u zadacima gdje je očuvanje strukturalnih detalja ključno.

Ovaj projekt koristi cGAN arhitekturu s U-Net generatorom i konvolucijskom mrežom kao diskriminatorom kako bi postigao visoku kvalitetu obojenih slika. Ovaj pristup omogućuje iskorištanje prednosti suvremenih tehnika dubokog učenja za postizanje vrhunskih rezultata u bojanju crno-bijelih slika.

III. OPIS SKUPA PODATAKA

U ovome radu koristili smo skup podataka COCO 2017. COCO (Common Objects in Context) 2017 skup je podataka koji je široko korišten u području računalnog vida. Dizajniran je s ciljem omogućavanja unapređenja razvoja algoritama za prepoznavanje objekata, segmentaciju, detekciju i opisivanje slika.

COCO 2017 sadrži više od 200 000 primjera. U našem radu, za učenje smo koristili 118 287 primjera, dok je za testiranje korišteno 40 670 primjera. Slike iz skupa podataka prikupljene su iz različitih scenarija svakodnevnog života - cilj je prikazivanje objekata u njihovoj prirodnoj okolini tj. kontekstu. Skup podataka sadrži označe za 80 različitih razreda objekata, uključujući ljude, životinje, vozila, kućanske predmete i brojne druge.



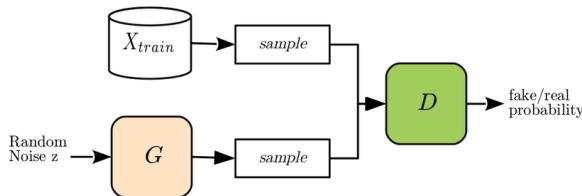
Slika 1: Primjer slika iz skupa COCO 2017. Preuzeto iz [4].

COCO 2017 ne koristi se samo za klasifikaciju pa stoga sadrži i neke pomoćne označe. Među njima su označe granica objekata tj. okviri (engl. *bounding boxes*), kao i precizne maske segmentacije za svaki objekt - ovo omogućava učenje modela za segmentaciju. Dodatno, skup podataka sadrži i označe ključnih točaka za ljudsko tijelo, uključujući označe ramena, laktova te koljena. Svaka slika ima više tekstualnih

opisa, što je korisno za zadatku opisivanja slike, kao i učenje jezičnih modela vezanih uz slike.

IV. ARHITEKTURA CGAN

Generativne suparničke mreže [5] razvio je istraživački tim pod vodstvom Iana Goodfellowa 2014. godine. Ova arhitektura sastoji se od dvije mreže koje rade jedna protiv druge (suparnički). Mreža generator na temelju nasumičnog vektora (najčešće uzorkovanog iz normalne razdiobe) treba generirati sliku koja bi idealno odgovarala distribuciji stvarnih podataka. S druge strane, diskriminator za slike na ulazu treba identificirati radi li se o stvarnoj ili umjetnoj (generiranoj) slici.

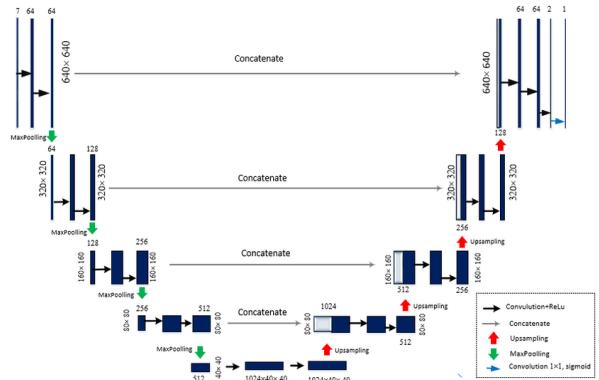


Slika 2: Arhitektura GAN. Preuzeto iz [6].

Tijekom učenja, diskriminator želi maksimizirati svoju pouzdanost klasifikacije - idealno, diskriminator bi za stvarne slike na izlazu dao vrijednost 1, dok bi za umjetne slike dao vrijednost 0. Istovremeno, generator želi generirati slike za koje će diskriminator što pouzdanije reći da su stvarne. Vidimo da je ovo igra s nultim zbrojem (engl. zero-sum game).

Za generator smo koristili arhitekturu zvanu U-Net [7]. U-Net je konvolucijska neuronska mreža koja je prvo bitno razvijena za zadatke segmentacije medicinskih slika, ali se s vremenom pokazala vrlo učinkovitom i u različitim drugim zadacima računalnog vida, uključujući segmentaciju, translaciju sa slike na sliku i bojanje slika. Arhitektura se sastoji od dva glavna dijela: kodera (downsampling path) i dekodera (upsampling path), koje zajedno formiraju oblik slova "U". Koder i dekoder oblikovali smo kao zasebne blokove u našem konačnom modelu.

Dodatno, arhitektura U-Net uključuje preskočne veze (engl. *skip connections*) između odgovarajućih slojeva kodera i dekodera. One omogućuju prijenos niskorazinskih značajki, čime se poboljšava točnost i preciznost rekonstrukcije slike. Preskočne veze ostvaruju se konkateniranjem značajki iz odgovarajućeg sloja kodera sa značajkama iz odgovarajućeg sloja dekodera.



Slika 3: Arhitektura U-Net. Preuzeto iz [8].

Koder se sastoji od niza konvolucijskih blokova. Svaki od tih blokova sadrži dva konvolucijska sloja s aktivacijom ReLU, kao i sloj sažimanja maksimumom (engl. *max-pooling*) za smanjenje dimenzija i povećanje receptivnog polja. Svaki blok smanjuje prostorne dimenzije i povećava broj kanala, što omogućuje modelu da uči sve složenije značajke.

Dekoder se također sastoji od niza konvolucijskih blokova. Ipak, kako bi postigli povećanje prostornih dimenzija, u dekoderskim blokovima koristi se transponirana konvolucija. Dodatno, koriste se i slojevi normalizacije po grupi (engl. *batch normalization*). Kao aktivacijska funkcija, u dekoderu se koristi propusna zglobnica (engl. *Leaky ReLU*). Osim transponirane konvolucije, za povećanje prostornih dimenzija se može koristiti i interpolacija - npr. bilinearna interpolacija. Svaki blok povećava prostorne dimenzije i smanjuje broj kanala, obnavljajući detalje slike.

Uvjetne generativne suparničke mreže (engl. *Conditional Generative Adversarial Networks* - cGAN) napredna su varijanta GAN-ova koja uvodi dodatne informacije u proces generiranja, ali i klasificiranja podataka. Ove mreže omogućuju generiranje uzorka uvjetovanih određenim unosom, kao što su slike, oznake razreda, tekstualni opisi ili druge vrste podataka. cGAN-ovi su vrlo korisni za zadatke gdje postoji potreba za kontroliranim generiranjem primjeraka poput bojanja slika, translacije sa slike na sliku i generiranja slika prema tekstualnim opisima. U našem radu, kao uvjetni vektor na ulaz generatora dovodimo crno-bijelu (engl. *grayscale*) sliku.

V. MJERE DOBROTE

Kada govorimo o evaluaciji performansi generativnih modела, postoje dva pristupa. Prvi od pristupa je vizualna procjena kvalitete - generirane slike ocjenjuju ljudi. Pošto je ovaj pristup veoma subjektivan, uvedene su brojne mjere dobrote tj. kvantitativne metrike. Ako je u pitanju zadatak poput bojanja slike gdje imamo očekivani izlaz, mogu se koristiti mjere poput srednje kvadratne pogreške (engl. *Mean Squared Error* - MSE) i mjere indeksa strukturalne sličnosti (engl. *Structural Similarity Index Measure* - SSIM). U našem radu koristili smo mjere IS (engl. *Inception Score*) i FID (engl. Fréchet Inception Distance).

IS [9] je mjera dobrote koja se koristi za procjenu kvalitete slika generiranih generativnim modelima, posebno GAN-ovima. Mjera koristi prednaučenu mrežu Inception v3 [10] za izračunavanje rezultata, a fokus mjere su dva ključna aspekta generiranih slika: kvaliteta i raznolikost. Konkretno, kvaliteta generiranih slika manifestira se visokom pouzdanošću klasifikacije mreže Inception v3 za pojedine slike. Drugim riječima, za svaku generiranu sliku, model bi na izlazu trebao dati distribuciju vjerojatnosti koja je fokusirana na jedan razred. Oba aspekta možemo kvantificirati koristeći KL-divergenciju izračunatu između distribucija vjerojatnosti za pojedine slike te distribucije marginaliziranih vjerojatnosti za sve razrede.

$$IS(G) = \exp \left(\mathbb{E}_{\mathbf{x} \sim p_g} D_{KL}(p(y|\mathbf{x}) \| p(y)) \right)$$

Slika 4: Formula za izračun IS-a.

FID [11] je mjera dobrote koja se također koristi za procjenu kvalitete generiranih slika. Kroz godine, ova mjera postala je standard u evaluaciji generativnih modela, posebno GAN-ova. FID mjeri sličnost između distribucija stvarnih i generiranih slika. Konkretno, za izračun mjere FID koristi se Fréchetova udaljenost između dvije multivarijatne normalne distribucije. Pritom distribucije modeliraju značajke iz posljednjeg sloja mreže Inception v3 - jedna distribucija modelira značajke za stvarne slike, dok druga distribucija modelira značajke za generirane slike.

$$FID = \|\mu_r - \mu_g\|^2 + T_r(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

Slika 5: Formula za izračun FID-a.

VI. EKSPERIMENTALNI REZULTATI

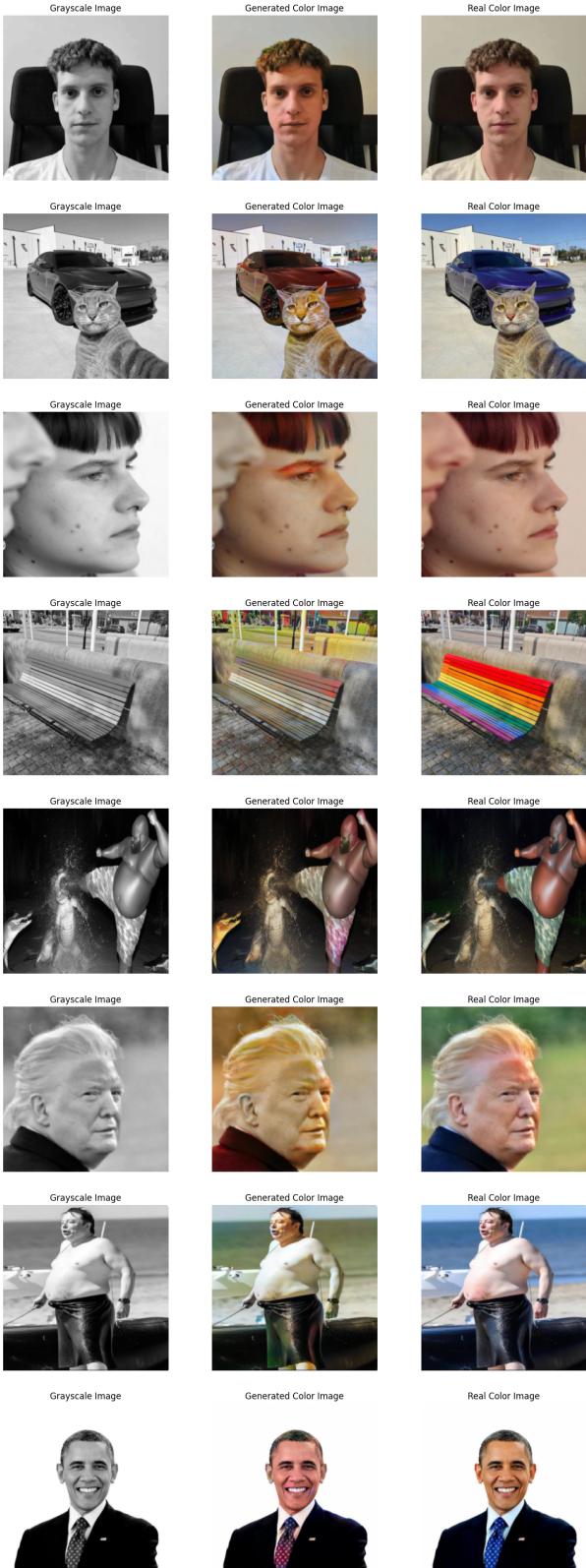
Slike za učenje modela na početku smo pretvorili u CIELAB prostor boja. Generator za zadalu crno-bijelu sliku (L komponentu originalne obojene slike) treba generirati komponente A i B. Na temelju generiranih A i B komponenti, stvara se obojena slika u CIELAB prostoru boja koja se prosljeđuje diskriminatoru.

Model smo učili 200 epoha. Za gubitak diskriminatora koristili gubitak unakrsne entropije, a za gubitak generatora koristili smo gubitak diskriminatora, kao i rekonstrukcijski gubitak (L1 udaljenost između umjetne i stvarne slike). Kao optimizator smo za obje mreže koristili optimizator Adam s početnom stopom učenja $2 \cdot 10^{-4}$. Nakon učenja modela, generator smo koristili za bojenje crno-bijelih slika.

Usporedno su prikazane: crno-bijela slika, slika obojena modelom cGAN i konačno stvarna slika u boji. Na slici 6 prikazan je izbor od nasumičnih 8 slika iz skupa COCO 2017 koje je model vidio pri učenju. Osim ovoga, prikazan je i izbor od 8 nasumičnih slika iz privatnih galerija ili s interneta koje model nije vidio pri učenju, na slici 7.



Slika 6: Primjer obojenih slika iz skupa COCO 2017



Slika 7: Primjer obojenih vlastitih slika

Na temelju ovog uskog izbora slika, može se primijetiti da model mnogo bolje radi na slikama koje prikazuju prirodne scene i sadrže nijanse smeđe i zelene. Naš model dobro radi

i na ljudskoj koži. Ovo možemo pripisati skupu podataka na kojemu je model učen - COCO 2017 sadrži velik broj slika na kojima su prikazani ljudi, kao i prirodne scene.

Osim provjere kvalitete obojenja ljudskim okom (koja je u ovakvom slučaju vrlo dobra mjera), potrebno je primijeniti objektivne matematičke mjere na izlaz modela. Korištene su, ranije spomenute, mjere IS (engl. *Inception Score*) i FID (engl. *Fréchet Inception Distance*). Rezultati su prikazani u tablici I.

Tablica I: Mjere dobrote modela

stvarne slike		obojene slike		
IS μ	IS σ	IS μ	IS σ	FID
5.2293	1.2776	5.1197	1.2776	8.5348

Za stvarne slike dobivena je mjera IS iznosa 5.2293 sa standardnom devijacijom od 1.2776. Za obojene slike mjera IS iznosi 5.1197 sa standardnom devijacijom od 1.2776. Kada govorimo o zadatku bojenja slika, cilj je da je mjera IS što sličnija za obojene slike i stvarne slike. Dodatno, bolje je da je mjera veća uz što manju standardnu devijaciju - veći iznos mjere IS govori nam da su slike kvalitetne i raznovrsne. Ovaj model ima malu razliku između mjere IS za stvarne slike i obojene slike, svega 0.1096.

Za naš model, dobivena je mjera FID iznosa 8.5348. Idealan iznos mjere FID je 0 - ovo označava da ne postoji razlika u distribuciji stvarnih i generiranih tj obojenih slika. Koliko je ovo uspješno može se vidjeti tek usporedbom s nekim drugim modelima.

VII. USPOREDBA S POSTOJEĆIM PRISTUPIMA

Za usporedbu, odabranu su sljedeća tri modela:

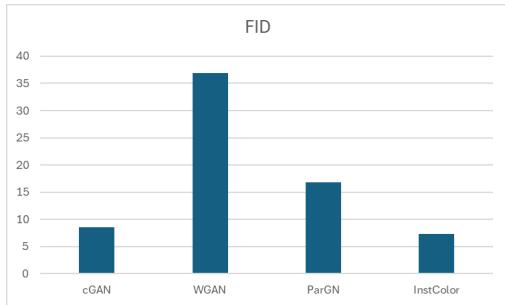
- WGAN - model Pix2pix uz Wasserstein GAN [12]
- ParGN - paralelni model temeljen na GAN-ovima [13]
- InstColor - Instance-aware Image Colorization [14]

Njihove mjere dobrote IS i FID usporedno su prikazane u tablici II. Nažalost, mjera IS nije izračunata u modelima ParGN [13] i InstColor [14]. Dodatno, mjera FID prikazana je stupčastim dijagramom 8, a mjera IS prikazana je za dva modela kao graf normalnih distribucija 9.

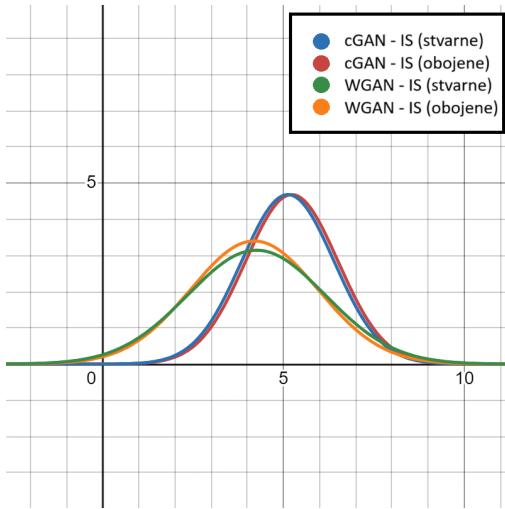
Tablica II: Usporedba mjera dobrote modela

Model	stvarne slike		obojene slike		FID
	IS μ	IS σ	IS μ	IS σ	
cGAN	5.2293	1.2776	5.1197	1.2776	8.5348
WGAN [12]	4.1913	1.7591	4.2828	1.9007	36.9385
ParGN [13]	-	-	-	-	16.8273
InstColor [14]	-	-	-	-	7.36

Kao što možemo vidjeti na prikazanom stupčastom dijagramu 8, naš model (cGAN) ima značajno bolji iznos mjere FID od modela WGAN [12] i ParGN [13], ali ne toliko dobar kao model InstColor [14]. To bi mogao biti rezultat ekstrakcije značajki objekata iz slika koje provodi model InstColor [14].



Slika 8: Stupčasti dijagram FID-ova modela



Slika 9: Graf IS-ova modela

Kao što možemo vidjeti na grafu 9, naš model (cGAN) ima veći iznos mjere IS od modela WGAN [12], uz manju standardnu devijaciju. Prema tome, naše obojene slike sličnije su stvarnim slikama u usporedbi s obojenim slikama dobivenim modelom WGAN [12]. Model WGAN [12] ima manju razliku središnje vrijednosti između stvarnih i obojenih slika iznosa 0.0915, dok model cGAN ima razliku iznosa 0.1096. Ipak, može se reći da cGAN ima bolji iznos mjere IS zbog znatno manje standardne devijacije, kao i općenito većeg iznosa.

VIII. ZAKLJUČAK

Razvijen je relativno jednostavan model, koji se može naučiti na osobnim računalima s grafičkim karticama. Bojenje je donekle uspješno, kao što se moglo vidjeti na slikama 6 i 7. Naučeni model vrlo dobro boji nijanse zelene i smeđe, kao i ljudsku kožu, ali ima problema sa šarenim površinama. Ova svojstva vezana su uz korišteni skup podataka za učenje.

Koristeći mjere dobrote FID i IS, vidljivo je da je model usporediv, ako ne i bolji, od sličnih modela za bojenje crno-bijelih slika. Prema mjeri dobrote FID, model je bolji od modela WGAN [12] i ParGN [13], a malo lošiji od modela InstColor [14] koji izlučuje značajke slika prije bojenja. Prema mjeri dobrote IS, model je bolji od modela WGAN [12].

Za daljnja istraživanja, može se koristiti više procesorske snage grafičkog procesora (GPU snage) kako bi se model

učio veći broj epoha. Povećanje skupa podataka također bi poboljšalo model - na primjer, model bi se mogao učiti na skupu podataka ImageNet [15]. Dodatno, mogu se isprobati modifikacije arhitekture, kao korištenje WGAN-a umjesto cGAN-a. Uz ovo, mogu se isprobati i potpuno drugačije arhitekture poput difuzijskih modela [16].

LITERATURA

- [1] S. Li, Q. Liu, H. Yuan, *et al.*, “Overview of scribbled-based colorization,” *Art and Design Review*, vol. 6, no. 04, p. 169, 2018.
- [2] R. Zhang, P. Isola, and A. A. Efros, “Colorful image colorization,” in *ECCV*, 2016.
- [3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” 2018.
- [4] S. Bourouis, I. Channoufi, R. Alroobaee, S. Rubaiee, M. Andejany, and N. Bouguila, “Color object segmentation and tracking using flexible statistical model and level-set,” *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5809–5831, 2021.
- [5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [6] J. Hayes, L. Melis, G. Danezis, and E. De Cristofaro, “Logan: Evaluating information leakage of generative models using generative adversarial networks,” *arXiv preprint arXiv:1705.07663*, vol. 18, 2017.
- [7] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015.
- [8] V. Raudonis, A. Paulauskaite-Taraseviciene, and K. Sutiene, “Fast multi-focus fusion based on deep learning for early-stage embryo image enhancement,” *Sensors*, vol. 21, no. 3, p. 863, 2021.
- [9] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” *Advances in neural information processing systems*, vol. 29, 2016.
- [10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [11] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” *Advances in neural information processing systems*, vol. 30, 2017.
- [12] S. Benhamdi, “Pix2pix : Image colorization with conditional wgan, version 31.” <https://www.kaggle.com/code/salimhammadi07/pix2pix-image-colorization-with-conditional-wgan>, 2023.
- [13] H. Kumar, A. Banerjee, S. Saurav, and S. Singh, “Paracolorizer: Realistic image colorization using parallel generative networks,” 2022.
- [14] J.-W. Su, H.-K. Chu, and J.-B. Huang, “Instance-aware image colorization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [16] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” in *International conference on machine learning*, pp. 2256–2265, PMLR, 2015.
- [17] N. Semary, *Image Coloring Techniques and Applications*. PhD thesis, 01 2011.