Upravljivo nadopunjavanje obrisanih regija slika lica

Tomislav Ćosić *FER Zagreb* Luka Družijanić *FER Zagreb* Renato Jurišić *FER Zagreb* Marko Kremer *FER Zagreb* Anto Matanović *FER Zagreb* Josip Srzić *FER Zagreb*

***Sažetak*—U ovom radu istražujemo primjenu generativnih modela u zadacima nadopunjavanja slika. Fokusiramo se na dva zadatka: općenito nadopunjavanje slika te upravljivo nadopunjavanje s mogućnošću promjene atributa. Za potrebe tih zadataka smo istrenirali cVAE model na skupu CelebA te ga fino ugodili koristeći suparnički gubitak iz GAN-ova. Takva kombinacija se pokazala povoljnom jer omogućuje generiranje oštrih slika s puno detalja uz stabilno treniranje. U radu smo detaljno opisali korištenu arhitekturu i metodologiju treniranja. Rezultati pokazuju uspješnost modela u navedenim zadacima, s naglaskom na modele koji su specijalizirani za svoj zadatak.**

I. Uvod

Razvoj generativnih modela posljednjih godina doveo je do njihove upotrebe u raznim zadacima. U ovom radu, tri su zadatka koje smo pokušali riješiti:

1) Nadopunjavanje (engl. *inpainting*) - na danoj slici potrebno je uvjerljivo nadopuniti regiju koja nedostaje
2) Upravljivo nadopunjavanje - na danoj slici potrebno je uvjerljivo nadopuniti regiju koja nedostaje uz opciju mijenjanja atributa (primjerice, obrisati regiju usta na slici lica te zatražiti da model rekonstruira usta sa ili bez osmijeha)

Nadopunjavanje je koristan alat u slučaju oštećenih, nezadovoljavajućih ili prekrivenih slika. Umjesto da ponovno slikamo istu sliku u nadi da popravimo mane na njoj, loše dijelove možemo sami izbrisati i nadopuniti generativnim modelima kako bismo popravili sliku.

Ekperimente radimo na CelebA skupu podataka sa dva popularna modela u ovoj domeni: uvjetovani varijacijski autoenkoder [4] i ugađanje (engl. *fine-tuning*) parametara pomoću suparničkog gubitka [2]. Model cVAE omogućava učenje bitnih značajki, dok GAN ugađanje pruža mreži učenje sitnih detalja.

Implementacija čitave mreže kao i naučene težine dostupne su na sljedećoj poveznici: https://github.com/LukaD00/cvae-image-inpainting.

II. Kratki pregled literature

Rane metode nadopunjavanja slike oslanjale su se na ručno izrađene značajke i plitke modele. Međutim, novije metode pokazale su da se duboko učenje može koristiti za postizanje stanja tehnike (engl. state-of-the-art) u području nadopunjavanja slika. Među metodama dubokog učenja, modeli kao što su VAE i GAN pokazali su se naročito uspješnima.

Yeh et al. [5] su pomoću DCGAN-a rekonstruirali velike dijelove slike na temelju konteksta okruženih piksela. Svoj rad najviše približuju Pathak et al. [3] kontekst enkoderu (CE). Pomoću maske se označe nedostajuća područja i onda se trenira neuronska mreža kako bi enkodirala kontekst i predvidjela nedostajuće dijelove. Prednost pristupa Yeh et al. [5] je da nije potrebna maska prilikom treniranja. Njihov model je najsličniji našem general inpainter-u, ali naš pristup rekonstruira manje pravokutnike.

Bao et al. [1], poput nas, kombiniraju arhitekturu uvjetovanih varijacijskih autoenkodera (cVAE) s arhitekturom GAN-ova kako bi generirane slike bile vizualno detaljnije i oštrije. Glavna razlika u arhitekturi u odnosu na naš pristup je korištenje klasifikatorske mreže na izlazu koja pomaže pri uvjetovanju modela s odabranim razredima. Takvu arhitekturu, kao i mi, treniraju s kraja na kraj (*engl. end-to-end*. Dodatno, uvode novi gubitak za učenje generatora koji ga tjera da minimizira očekivanu L2 udaljenost srednje vrijednosti generiranih i stvarnih podataka u prostoru značajki. Pokazali su da taj pristup dovodi do stabilnijeg učenja GAN-a. Također, u jedan od gubitaka koji se odnose na generator uz klasični rekonstrukcijski član uvode i L2 udaljenost između značajki ulazne slike i njene rekonstrukcije. Autori su pokazali da na taj način dobiju veću raznolikost generiranih primjera. Za izvlačenje značajki koriste jedan od slojeva iz mreže diskriminatora.

III. Metoda

Zadatci su ostvareni pomoću VAE modela. VAE je generativni model koji uči probabilističku distribuciju podataka u latentnom prostoru. Prvo, enkoder kompresira sliku u sažetu, apstraktnu reprezentaciju z koja sadrži sve bitne informacije kako bi se ulazna slika mogla rekonstruirati. Zatim, dekoder iz te reprezentacije rekonstruira cijelu sliku. Ideja je da VAE enkoder pronađe generalno područje u latentom prostoru gdje se slika nalazi, a onda će uzorkovanje iz tog područja rezultirati sličnim slikama. Uzorkovanjem smo postigli da imamo varijaciju kada nadopunjavamo obrisane regije lica.

Kako bismo ostvarili nadopunjavanje, nije potrebno ništa mijenjati u modelu. VAE model bi trebao moći naučiti ignorirati obrisane regije slike i uvjerljivo ih nadopuniti. Jedina izmjena je u postupku učenja, gdje modelu tijekom učenja predajemo slike s obrisanim regijama, dok kvalitetu rekonstrukcije uspoređujemo sa originalnom slikom.

Međutim, kako bismo ostvarili *upravljivo* nadopunjavanje ili rekonstrukciju, potrebno je uvesti oznake u model, čime dobivamo cVAE model. Oznaku dodajemo samo u dekoderu. Najlakši način da se doda informacija o oznaci u rekonstrukciju je da zalijepimo njeno ugrađenje (engl. embedding) na uzorak. Dakle, na uzorak koji se dobije nakon uzorkovanja latentnog prostora se još zalijepi ugrađenje koje će usmjeravati

rekonstrukciju i onda se taj cijeli vektor predaje dekoderu. Ugrađenje oznake može biti fiksno, ali je preferirano to ugrađenje učiti skupa s drugim parametrima modela kako bi model pronašao najbolju reprezentaciju oznake koja mu odgovara. Oznaka služi kako bi usmjerila rekonstrukciju, npr. ako sa slike nečijeg lica obrišemo usta, onda pomoću oznake možemo odrediti želimo li da se osoba na slici nakon rekonstrukcije smije ili je ozbiljna. Cijela je arhitektura prisutna na slici (Slika. 1).

Običan VAE model generira mutne slike, što se dogodilo i u ovom slučaju. Zato smo cjelokupni model trenirali u dva navrata. Prvi put kao običan VAE model koji pokušava rekonstruirati ulaznu sliku, a zatim smo u drugom navratu dodali suparnički gubitak kako bismo ispravili mutnu teksturu slika.

A. Predtreniranje modela

Model predtreniramo kao običan VAE model tako da na ulaz dostavljamo slike koje želimo da on rekonstruira na izlazu. Dodatno, generatoru dostavljamo i oznaku koja usmjerava rekonstrukciju. Ta oznaka, tijekom treniranja, će biti atribut koji je prisutan na slici. Npr. ako treniramo VAE na slikama lica za koje želimo da on usmjereno rekonstruira osmjeh, onda ćemo generatoru kao oznaku dati da li se osoba na ulazu smiješi ili je ozbiljna. Na taj način će model povezati da kada dobije oznaku za osmijeh treba generirati osobu koja se smiješi, a kada dobije oznaku za ozbiljno lice treba generirati osobu bez osmjeha. (Slika. 1).

Kao i u standardnom VAE treniranju, imamo 2 gubitka. Prvi gubitak je gubitak rekonstrukcije:

$$L_{BCE} = \sum_{i=1}^{N} \left[\mathbf{X} \cdot \ln\left(\hat{\mathbf{X}}\right) + (1 - \mathbf{X}) \cdot \ln\left(1 - \hat{\mathbf{X}}\right) \right], \quad (1)$$

gdje je $\mathbf{X}$ originalna slika, a $\hat{\mathbf{X}}$ rekonstrukcija slike koju je model dao na izlazu. Gubitak rekonstrukcije mjeri koliko rekonstrukcija odstupa od originala. Za gubitak rekonstrukcije se koristi gubitak binarne unakrsne entropije, a ne kvadratna pogreška jer u zadnjem sloju imamo sigmoidu. Kvadratna pogreška u kombinaciji sa sigmoidom može dovesti do zasićenja, zbog čega gradijenti ne bi propagirali nazad u mrežu. Kako bismo to izbjegli, koristimo binarnu unakrsnu entropiju koja ima logaritam koji poništava potenciju e^x u sigmoidi, pa ne dolazi do preuranjenog zasićenja.

Drugi gubitak je divergencija distribucije latentnog prostora od standardne normalne razdiobe:

$$L_{KL} = D_{KL}(\mathcal{N}(\hat{\mu}, \hat{\sigma}^2) \,||\, \mathcal{N}(0,1)) = \\ -1 - \ln\hat{\sigma} + \hat{\sigma} + \hat{\mu}^2, \quad (2)$$

gdje su ($\hat{\mu}$ i $\hat{\sigma}$ parametri distribucije latentni varijable z koju enkoder daje na svom izlazu. Konačan gubitak je:

$$L = L_{BCE} + 0.5 \cdot L_{KL} \quad (3)$$

Za nadopunjavanje (zadatci 1 i 2), razlika koju radimo u odnosu na standardni način treniranja VAE modela je da iz slike na ulazu tijekom treniranja nasumično obrišemo jedan pravokutnik. Ovo približuje treniranje modela načinu na koji će raditi inferenciju. Na ulaz će dobijati slike s nedostajućom regijom i od njega se traži da rekonstruira kompletnu sliku.

B. Fino ugađanje modela

Standardni VAE generira mutne slike. Kako bi penalizirali takvo ponašanje modela, uvodimo suparnički gubitak. Suparnički gubitak je druga konvolucija mreža koja se zove diskriminator. Uloga diskriminatora je da nauči razlike između pravih i generiranih slika i propagira tu informaciju do generatora kako bi se on mogao poboljšati (Slika 2).

Razlog zašto odmah nije uveden suparnički gubitak prilikom treniranja u prvom navratu je jer je treniranje GAN modela relativno nestabilno. Na ovaj način, kada model treniramo u dva navrata, generator već u prvom navratu nauči generirati slike, ali su one mutne. Onda je u drugom navratu potrebno samo fino ugodit generator da izbjegne mutne slike. Konvolucijske mreže su dobre u prepoznavanju teksture, pa će jednostavni diskriminator moći jednostavno uočiti razliku i propagirati tu informaciju generatoru.

Tijekom finog ugađanja je i dalje prisutan gubitak rekonstrukcije i gubitak divergencije distribucije, ali se dodaje suparnički gubitak. Može se prioritizirati između ta tri gubitka tako da ih pomnožimo relativnim težinama. Tako se u ovom navratu veća težina predaje suparničkom gubitku.

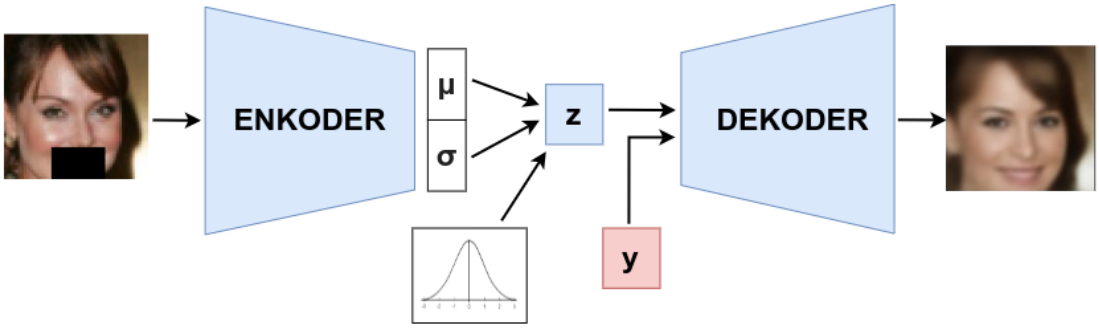
Važno je napomenuti da prije nego što krenemo iterativno trenirati diskriminator i generator, je potrebno prvo par iteracija trenirati samo diskriminator kako bi se on zagrijao i došao do razine znanja koju trenutno ima generator. Ako nemamo taj period zagrijavanja, onda će diskriminator na početku generatoru slati beskorisne gradijente koji će uništiti znanje koje je generator nakupio u prvom navratu treniranja.

IV. Eksperimentalni rezultati

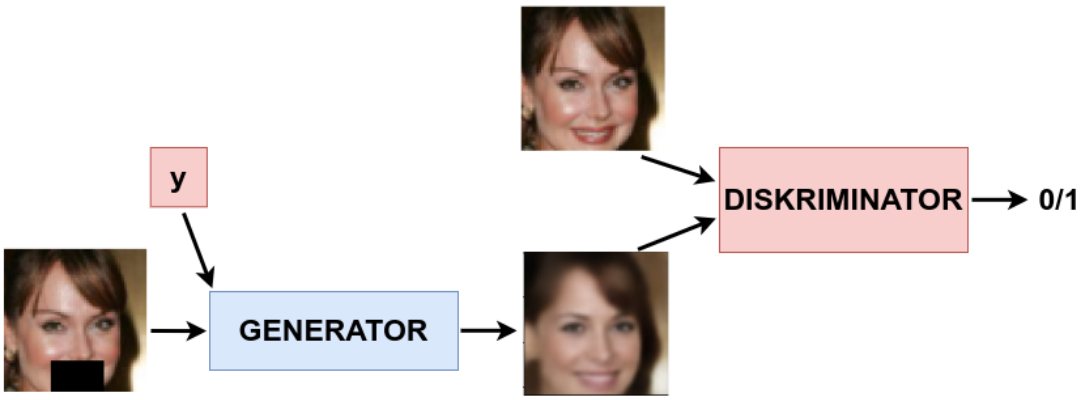
Eksperimente provodimo na CelebA skupu podataka, koji sadrži 202599 poravnatih (engl. *aligned*) slika lica. Zbog ograničenih računalnih resursa, slike su skalirane i odrezane na veličinu od 64x64 piksela. Za enkoder, dekoder i diskriminator korišteni su duboki modeli sa četiri sloja konvolucije, batchnorma i aktivacije. Pokušali smo dodati i slojeve pažnje u modele, no nismo uočili značajan napredak. Naš konačan cVAE model ima 13 milijuna parametara. Treniranje se provodi Adam optimizatorom sa 0.01 stopom učenja i veličinom grupe 128. Na našim računalima, jedna epoha traje 5-10min. Treniranje smo provodili do konvergencije, što je trajalo od 5 do 30 epoha.

Naučili smo ukupno 4 modela:

1) *general-inpainter* - Nadopunjuje "male" pravokutnike na slici, bez atributa
2) *general-inpainter-big* - Nadopunjuje "velike" pravokutnike na slici, bez atributa
3) *smiling-inpainter* - Nadopunjuje pravokutnik na ustima, s opcijom da nadopuni usta sa ili bez osmijeha



Slika 1. Arhitektura cijelog modela. Sastoji se od VAE enkodera i VAE dekodera kojemu je dodano ugrađenje oznake kako bi se usmjerila rekonstrukcija.



Slika 2. Treniranje enkodera kao generatora u GAN-u.

4) *glasses-remover* - Rekonstruira sliku brišući naočale s nje

Kvalitativni rezultati naših modela vidljivi su na slikama 3-6.

Veći pravokutnici su se pokazali težim zadatkom za nadopunit. Generalno, uspješnost modela značajno ovisi o načinu brisanja pravokutnika. Primjerice, *smiling-inpainter* je imao lošije performanse kada je bio treniran na nasumično postavljenim pravokutnicima. Dodatno, *smiling-inpainter* je bolji u nadopunjavanju usta od *general-inpainter* modela. Model *general-inpainter* je jako loš na većim pravokutnicima, a *general-inpainter-big* ima dobre rezultate na malim pravokutnicima, ali lošije od *general-inpainter*.

V. Zaključak

Upravljivo nadopunjavanje korisna je primjena generativnih modela. U ovom radu demonstrirali smo uspješnost cVAE modela na CelebA skupu podataka. Pokazali smo da fino ugađanje modela sa suparničkim gubitkom pridonosi boljoj kvaliteti slika. Naši modeli daju dobre nadopune, no veći modeli bi mogli biti daleko uspješniji. Naši modeli dobro specijaliziraju za specifične zadatke, no za dobru općenitu nadopunu s više atributa potrebni su puno jači modeli.

Literatura

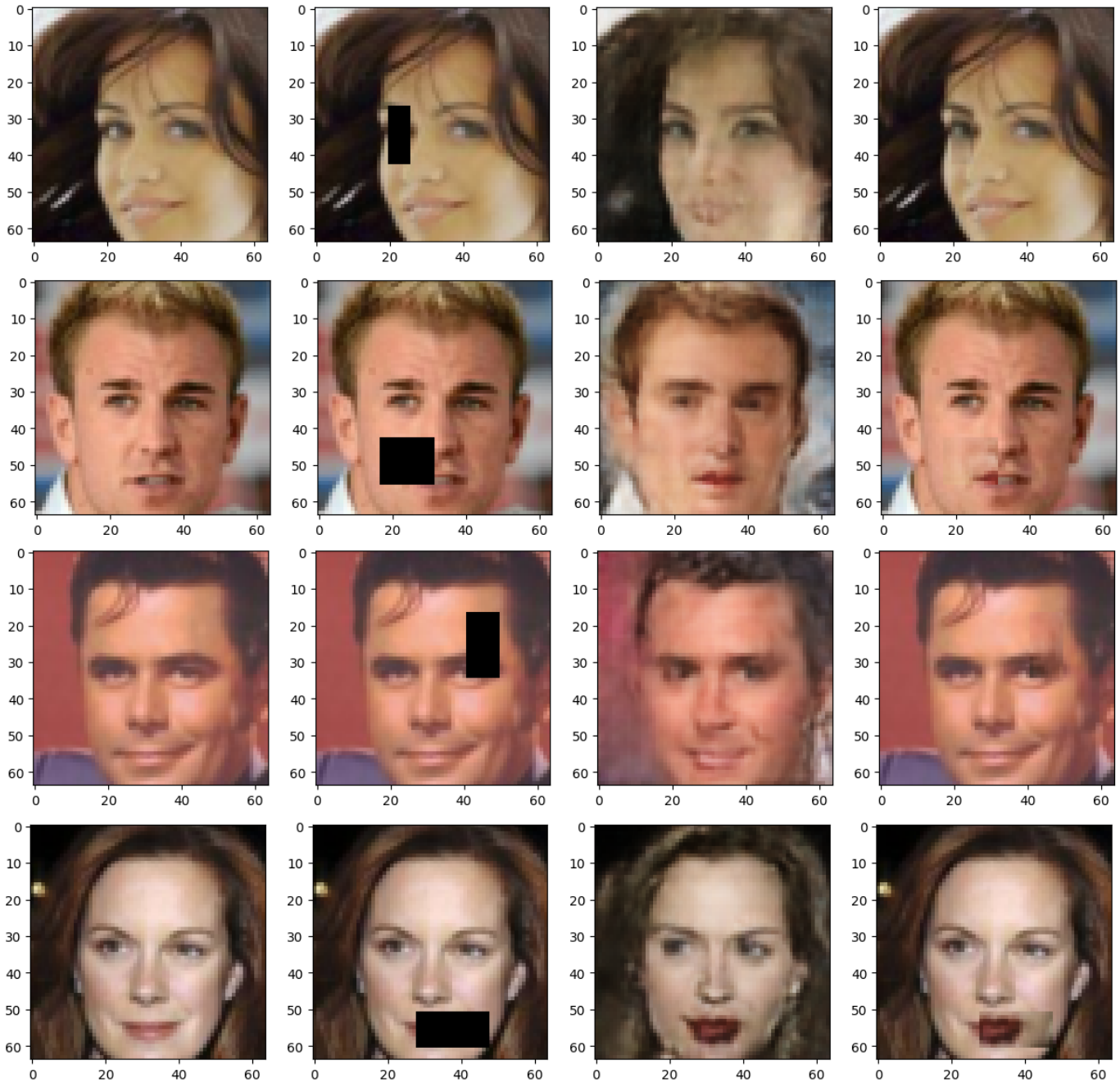
[1] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. Cvae-gan: Fine-grained image generation through asymmetric training. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

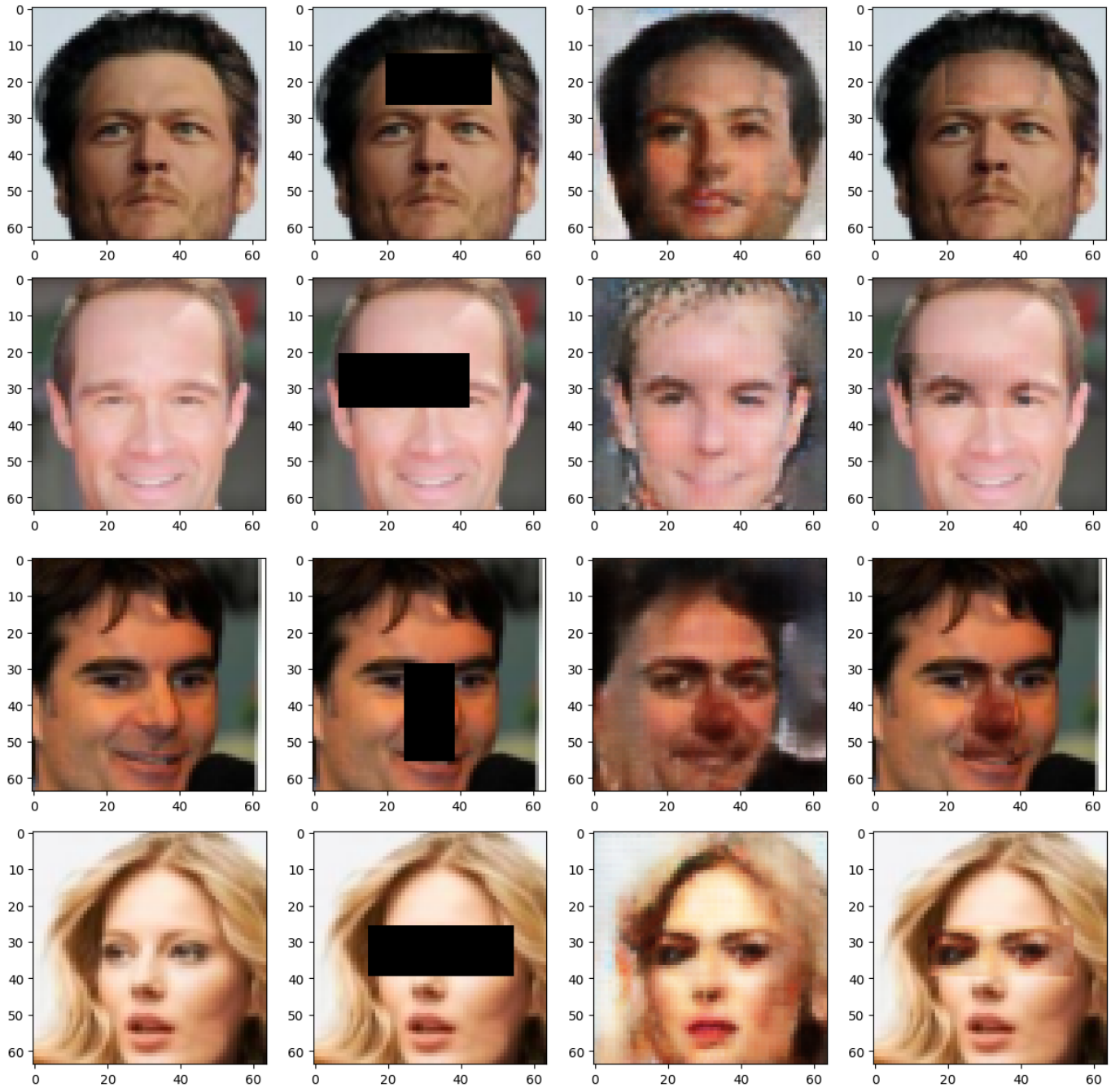
[3] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[4] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

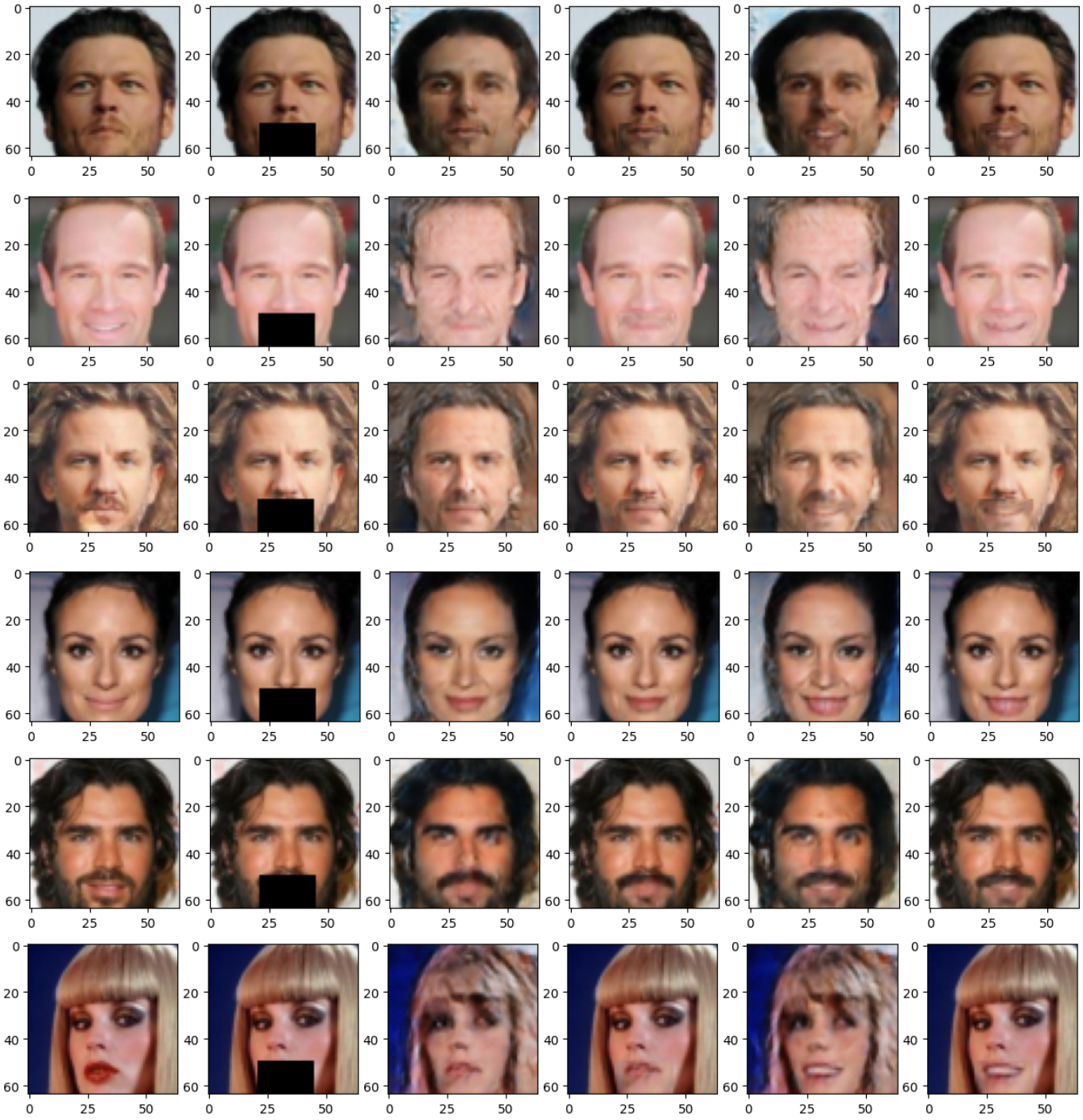
[5] Raymond A. Yeh, Chen Chen, Teck Yian Lim, Alexander G. Schwing, Mark Hasegawa-Johnson, and Minh N. Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.



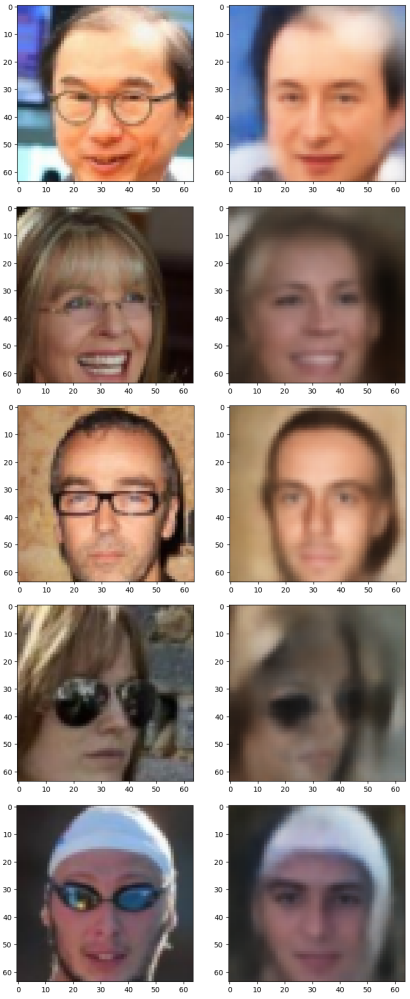
Slika 3. Rezultati *general-inpainter* modela. Slike prikazuju redom: original, original s uklonjenim pravokutnikom, rekonstrukcija modela, original s nadopunom modela na mjestu pravokutnika



Slika 4. Rezultati *general-inpainter-big* modela. Slike prikazuju redom: original, original s uklonjenim pravokutnikom, rekonstrukcija modela, original s nadopunom modela na mjestu pravokutnika



Slika 5. Rezultati *smiling-inpainter* modela. Slike prikazuju redom: original, original s uklonjenim pravokutnikom, rekonstrukcija modela bez osmijeha, original s nadopunom modela bez osmijeha na mjestu pravokutnika, rekonstrukcija modela sa osmijehom, original s nadopunom modela sa osmijehom na mjestu pravokutnika



Slika 6. Rezultati *glasses-remover* modela. Slike prikazuju redom: original, rekonstrukciju modela bez naočala