

Strukturni model za sintezo obrazov sorodstva z mehanizmi za nadzor videza

Martin Pernuš, Vitomir Štruc, Simon Dobrišek

Fakulteta za elektrotehniko, Univerza v Ljubljani

E-pošta: martin.pernus1@gmail.com

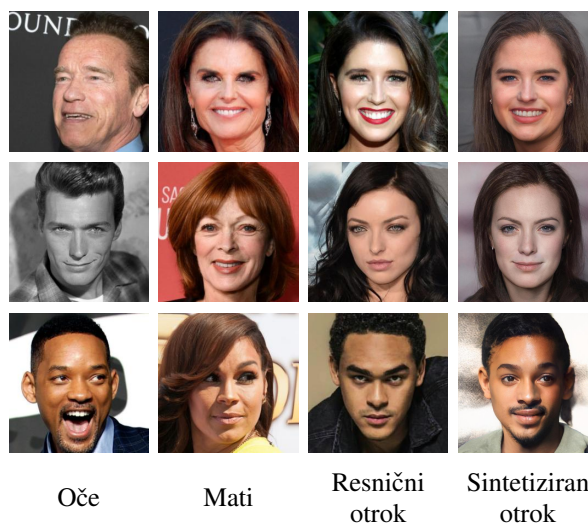
Povzetek

Sinteza obrazov sorodstva je vse bolj priljubljena tema znotraj skupnosti računalniškega vida, še posebej naloga napovedovanja videza otroka z uporabo starševskih slik. V tem članku predlagamo ChildNet, metodo za sintezo obrazov sorodstva, ki izkorišča generativne sposobnosti generativnih nasprotniških omrežij. ChildNet je oblikovan znotraj latentnega prostora GAN in je zmožen napovedati videz otroka, ki je zelo podoben resničnim otrokom staršev. Za zagotavljanje finega nadzora predlagamo modul za manipulacijo starosti in spola, ki omogoča natančno manipulacijo rezultata sinteze otroka. Poleg tega oblikujemo mehanizem za nadzorovanje prevladujoče starševske slike in variabilnosti slike na osnovi fiksne para staršev. ChildNet ovrednotimo v eksperimentih proti trem konkurenčnim modelom za sintezo obrazov sorodstva na dveh sorodstvenih podatkovnih zbirkah. Rezultati izkazujejo boljše zmogljivosti predlaganega modela ChildNet v metriki obrazne podobnosti sintetiziranih slik.

1 Uvod

Sinteza obrazov sorodstva je vznikajoče področje na področju računalniškega vida, ki zaobjema ustvarjanje slik obrazov, ki spominjajo na resnične sorodstvene osebe, ki jih ustvarimo na podlagi ene ali več vhodnih slik obrazov. Tehnike, ki so sposobne avtomatske sinteze obrazov sorodstva, imajo ogromen potencial na različnih področjih, kot so zabavna industrija, zgodovinske raziskave in forenzika, ki vključuje iskanje dolgo izgubljenih družinskih članov. Posledično so se raziskovalna prizadevanja nedavno začela osredotočati na sintezo obrazov sorodstva, kar je prineslo prepričljive rezultate [3, 1, 8].

Metode za sintezo obrazov sorodstva so nedavno doživele velik napredek, predvsem zahvaljujoč uspehu modelov za generiranje slik, ki uporabljajo generativna nasprotniška omrežja (GANs) [4]. Te metode izkoristijo nizkodimenzionalni latentni prostor GAN, tako da vhodne slike pretvorijo v latentne kode. Te kode se nato obdelujejo z namenskimi operacijami. Za pridobitev latentne kode iz slike se običajno uporablja pristop obrnjene GAN, ki izlušči latentno kodo, ki približno ustreza vhodni sliki, ko jo dekodira GAN. Z uporabo obrnjene GAN na starševskih slikah se pridobijo starševske kode, ki se nato obdelujejo z uporabo posebnih operacij na latentnih



Slika 1: Predstavljamo novo metodo za sintezo obrazov sorodstva, sposoben sintetizirati visokoločljivostno sliko otroka, ki močno spominja na resničnega otroka.

kodah, ki so običajno navdihnjene z genetskimi mehanizmi. Rezultat operacij je nova (otroška) latentna koda, ki jo obdelava GAN dekodler, ustreza sliki obraza, ki se lahko šteje za resničnega potomca staršev.

Ključni izziv pri izboljšanju zmogljivosti obstoječih modelov je povečanje kapacitete modela na način, ki izboljša podobnost sintetiziranega rezultata z videzom resničnega otroka. Poleg tega obstoječi modeli ne zagotavljajo ustrezne finega nadzora nad procesom sinteze, kot je nadzor nad raznolikostjo sintetiziranih slik obrazov in nadzor nad prevladujočo starševsko sliko.

V tem članku predstavljamo nov model z imenom ChildNet, ki dodatno izboljša zmogljivosti sinteze obrazov sorodstva in je sposoben generirati visoko ločljive vizualno prepričljive rezultate, kot je prikazano na sliki 1. Model se uči na celovit način, zato v implementaciji modela ni treba določiti nobenega vnaprej določenega zunanjega (genetskega) znanja. Poleg tega ChildNet predstavlja več tehnik za manipulacijo slik, da omogoči nadzor nad želenim videzom otroka. Na podlagi eksperimentov z dvema naboroma podatkov o sorodstvu ChildNet dosega večjo podobnost z resničnimi slikami otrok v primerjavi s konkurenčnimi metodami.

2 Sorodno delo

Generativna nasprotniška omrežja (angl. Generative Adversarial Network, GAN). Modeli GAN so postale eden najbolj priljubljenih modelov za ustvarjanje slik v zadnjih letih. Od njihovega nastanka [4], so se modeli GAN izboljšali v stilu arhitekture omrežja in postopkih učenja. Karras *et al.* [5] je predlagal ProGAN, ki je uporabil postopno učenje, kar je omogočilo ustvarjanje prepričljivih slik z megapiksel ločljivostjo. Model je bil nadaljnje izboljššan s predstavitvijo StyleGAN [6], katerega oblikovanje je bilo navdihneno z arhitekturo prenosa stila. Model je v naslednji iteraciji, StyleGAN2 [7], spremenil arhitekturno zasnovo, s čimer so se odstranili določeni vizualni artefakti na ustvarjenih slikah, obenem pa se je izboljšala tudi celotna zmogljivost sinteze.

Sinteza sorodstva. Sinteza sorodstva je relativno novo raziskovalno področje, kjer je najbolj priljubljena naloga sinteza videza otroka na podlagi starševskih slik. DNA-Net [3] je uporabil nevronske omrežje, kjer je kodirnik slik ustvaril značilnosti "genov", ki so se nato kombinirali glede na pravila izbire genov, preden so bile dekodirane v končno sliko otroka. Cui *et al.* [1] je predstavil model, ki je združeval starševske kode na podlagi genetskega biološkega znanja o dednih lastnostih, povezanih z velikostjo posameznih obraznih značilnosti. StyleDNA [8] je določil model kodirnik-dekodirnik, ki je nadzoroval starost in spol sintetizirane slike otroka, medtem, ko je bila strategija združevanja genov navdihnena z DNANet.

3 Metodologija

Pregled metode ChildNet. Cilj ChildNet-a je sintetizirati sliko otrokovega obraza iz dveh slik staršev. Dodatne informacije v obliki starosti in spola se lahko uporabijo za dodatno prilagoditev videza sintetiziranega otroka. ChildNet je razdeljen na dva glavna podmodula: Sorodstveni modul, ki določa latentno kodo otroka na podlagi latentnih kod staršev, in Modul za manipulacijo s starostjo in spolom, ki po potrebi spremeni latentno kodo otroka na podlagi danih informacij o starosti in spol. Latentna koda posameznega starša je pridobljena s prednaučenim modelom preslikave GAN-inverzije. Ko je končna latentna koda otroka sintetizirana, se prenese skozi model dekodirnika. Kot dekodirnik uporabljamo StyleGAN2 [7], ki je sposoben sintetizirati slike obraza z visoko ločljivostjo. Strukturna zasnova ChildNet-a omogoča neodvisno modeliranje različnih obraznih značilnosti pri otroku na podlagi podmnožice starševskih latentnih kod.

Modul sorodstva. Modul sorodstva je prvi podmodul ChildNet-a. Glede na vhodne slike staršev je glavni cilj tega modula sinteza videza otroka, ki ga je mogoče realistično obravnavati kot pravega potomca predloženih staršev. Zagotavlja, da je latentna koda otroka nekje blizu latentnega podprostora, kot je določeno z latentnimi kodami staršev. Poleg tega modul rahlo dodatno premakne napovedano latentno kodo. Ta premik, ali "mutacija", pomaga razčleniti kompleksne obrazne značilnosti, kodirane v starševskih latentnih kodah. Modul sorodstva je realiziran z dvema ločenima nevronska omrežjema; prvo

omrežje zagotavlja, da je latentna koda blizu starševskega latentnega podprostora, drugo omrežje pa napove latentni premik. Obe omrežji temeljita na večplastni polno-povezani arhitekturi.

Modul za manipulacijo starosti in spola. Drugi podmodul ChildNet-a je mehanizem za manipulacijo s starostjo in spolom, ki omogoča natančnejši nadzor nad sintetizirano sliko otroka. Modul za starost in spol sprejme sliko obraza (v obliki njene latentne kode), ciljno starost in ciljni spol kot vhodne podatke. Njegov cilj je sintetizirati sliko obraza, ki ohranja identiteto osebe, medtem ko prepričljivo manipulira z njihovo starostjo in spolom proti ciljnim vrednostim. Modul je realiziran v obliki večplastnega polno-povezanega nevronskega omrežja.

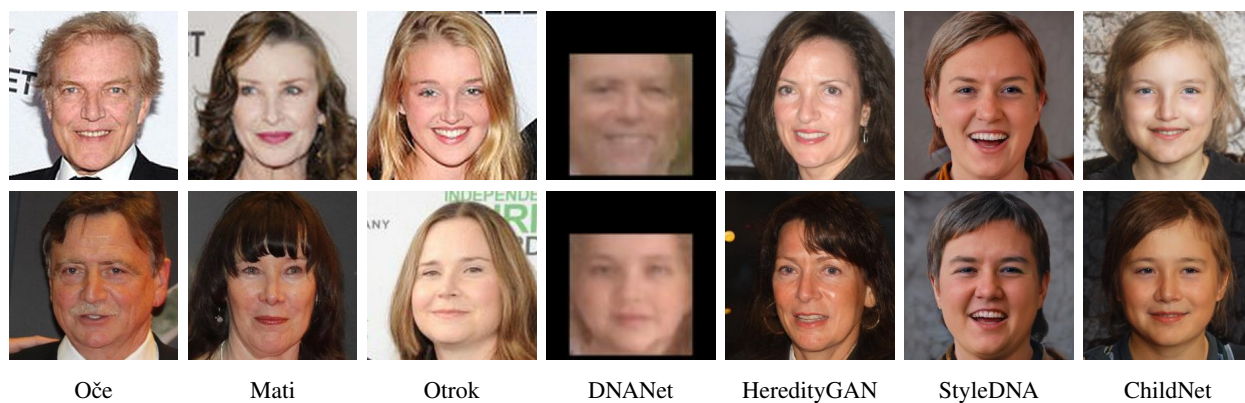
4 Eksperimenti

Eksperimentalni protokol ChildNet primerjamo z obstoječimi metodami z uporabo dveh podatkovnih zbirk. Prva podatkovna zbirka, Families in the Wild [9], se tradicionalno uporablja za učenje modelov za sorodstveno sintezo. Druga podatkovna zbirka je naša nova predlagana podatkovna zbirka Next of Kin, ki vsebuje 3690 visokorezolucijskih slik 161 družin, ki so ustreznejše za učenje modele sinteze. Model ChildNet ovrednotimo proti več konkurenčnim modelom: DNANet [3], HeredityGAN [1] in StyleDNA [8]. Določimo dve vrsti poskusov: poskus neomejene sinteze, kjer je napovedan videz otroka pogojen izključno s slikami staršev, in poskus pogojene sinteze, kjer testiranim modelom priskrbimo tudi informacije o starosti in spolu.

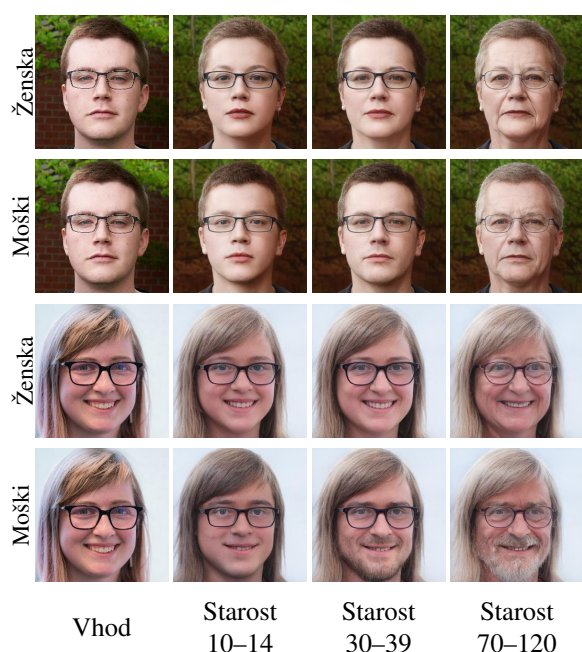
Vizualna analiza, nepogojeni eksperiment. Na sliki 2 primerjamo vizualne rezultate modelov za nalogo nepogojene sinteze otroka, torej brez dodatnih informacij o starosti in spolu. ChildNet v primerjavi z ostalimi metodami doseže najbolj vizualno privlačne rezultate in sintetizira slike, ki se jih lahko prepričljivo obravnava kot prave potomce vhodnih slik staršev.

Visualna analiza, pogojeni eksperiment. V pogojem eksperimentu ChildNet uporabi modul za manipulacijo starosti in spola. Primer delovanja modula je prikazan na sliki 3, kjer z modulom nadzorujemo spol in starost posameznega vhodnega obraza z manipulacijo latentne kode dane slike. Ko modul uporabimo skupaj z modulom sorodstva, je model ChildNet sposoben natančnejše sorodstvene sinteze, saj sintetiziran obraz otroka odraža podan spol in starost. Vizualni rezultati za nalogo pogojene sinteze otroka, torej sintezo z dodano informacijo o starosti in spolu otroka, so prikazani na sliki 4. Opazimo lahko, da tudi v tem primeru ChildNet daje v primerjavi z ostalimi modeli najboljše vizualne rezultate, ki so najbolj podobni resničnemu otroku.

Kvantitativna analiza. Za oceno točnosti sintetizirane identitete ocenjujemo podobnost identitete z uporabo modela za prepoznavanje obrazov ArcFace [2], merimo povprečno kosinusno podobnost vdelave med sintetizirano in resnično sliko otroka. Kvantitativni rezultati v obliki povprečne podobnosti ArcFace so prikazani v tabeli 1. ChildNet doseže najboljše rezultate v pogojnih in nepogojnih poskusih na obeh podatkovnih zbirkah, kar



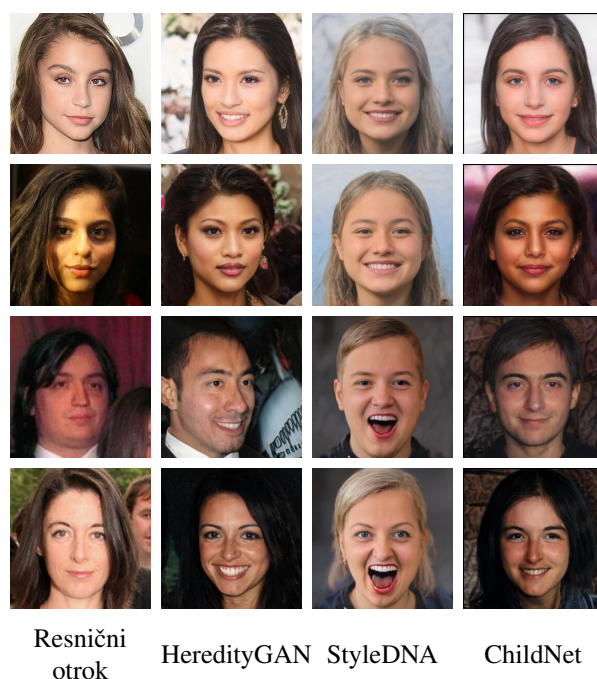
Slika 2: Prvi trije stolpci predstavljajo sliko očeta, slika matere in slika resničnega otroka, v tem vrstnem redu. Naslednji štirje stolpci prikazujejo nepogojen rezultat za vsak model.



Slika 3: Prikaz delovanja modula za manipulacijo starosti in spola.

Tabela 1: Povprečni rezultati podobnosti identitete, kot ovrednoteno z modelom ArcFace [2].

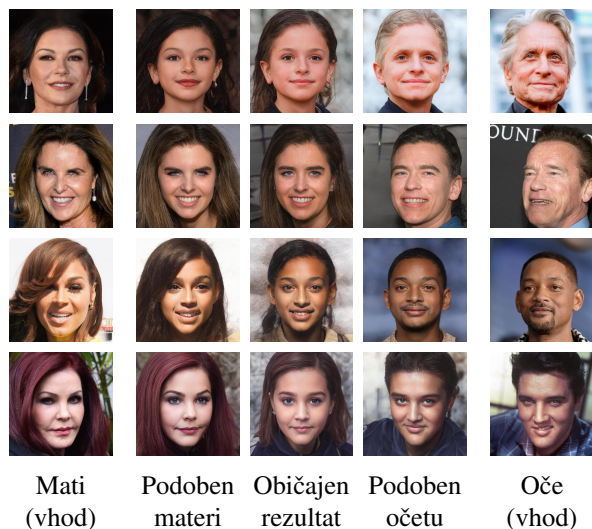
	Brezpogojni eksperiment	
	Families in the Wild	Next of Kin
DNANet [3]	0.058	-0.017
HeredityGAN [1]	0.113	0.094
StyleDNA [8]	0.123	0.153
ChildNet (naš)	0.280	0.205
	Pogojni eksperiment	
	Families in the Wild	Next of Kin
DNANet [3]	0.087	0.018
HeredityGAN [1]	0.145	0.146
StyleDNA [8]	0.237	0.161
ChildNet (naš)	0.299	0.221



Slika 4: Vizualni primeri pogojenega eksperimenta na podatkovni zbirki Next of Kin. Modelom poleg slik staršev predstavimo tudi informacijo o spolu in starosti za natančnejšo sintezo.



Slika 5: Prikaz mehanizma za nadzor slikovne variabilnosti.



Slika 6: Prikaz mehanizma dominantnega starša

kaže na sintezo otrok, ki bolje ustreza resničnemu videzu otroka.

Vzorčenje otrok in nadzor nad variabilnostjo. V tem poskusu analiziramo mehanizem variabilnosti slik ob fiksnih vhodnih slikah staršev. Predlagamo uporabo mehanizma izpuščanja [10] v modulu za sorodstvene vezi, da zagotovimo variabilnost videza otroka. Z uporabo mehanizma izpuščanja ChildNet lahko sintetizira variacije slik ob fiksnih vhodnih slikah staršev. Poleg tega lahko ChildNet z variiranjem stopnje izpuščanja nadzoruje variabilnost slik sintetiziranih rezultatov, kot je prikazano na sliki 5.

Nadzor dominantne slike staršev. Predlagamo mehanizem, ki nadzoruje videz otroka, da bolj spominja na enega od staršev. To je izvedeno z modificiranjem interpolacijskega preslikovanja modula za sorodstvene vezi. Vizualni primeri so prikazani na sliki 6, ki prikazuje, da naš mehanizem res omogoča sintezo bolj materinskega ali očetovskega videza, hkrati pa ohranja visoko kakovost slike.

5 Zaključek

V tem članku predstavljamo ChildNet, nov model za sintezo obrazov sorodnikov. ChildNet dosega najboljše rezultate glede na podobnost identitete, medtem ko kaže vsestranskost v smislu sintetiziranja več slik na vnos, nad-

zorom nad variabilnostjo sinteze in prilagajanje dominantne starševske slike.

Literatura

- [1] Xiao Cui, Wengang Zhou, Yang Hu, Weilun Wang, and Houqiang Li. Heredity-aware child face image generation with latent space disentanglement. *arXiv preprint arXiv:2108.11080*, 2021.
- [2] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.
- [3] Pengyu Gao, Siyu Xia, Joseph Robinson, Junkang Zhang, Chao Xia, Ming Shao, and Yun Fu. What will your child look like? dna-net: Age and gender aware kin face synthesizer. *arXiv preprint arXiv:1911.07014*, 2019.
- [4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems (NIPS)*, 2014.
- [5] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations (ICLR)*, 2018.
- [6] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410, 2019.
- [7] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Computer Vision and Pattern Recognition (CVPR)*, pages 8110–8119, 2020.
- [8] Che-Hsien Lin, Hung-Chun Chen, Li-Chen Cheng, Shu-Chuan Hsu, Jun-Cheng Chen, and Chih-Yu Wang. Styledna: A high-fidelity age and gender aware kinship face synthesizer. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE, 2021.
- [9] Joseph P Robinson, Ming Shao, Yue Wu, Hongfu Liu, Timothy Gillis, and Yun Fu. Visual kinship recognition of families in the wild. *IEEE Transactions on pattern analysis and machine intelligence*, 40(11):2624–2637, 2018.
- [10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.