

# Nadopunjavanje slike korištenjem difuzijskih modela

Martin Bakač, Mislav Đomlija, Ivan Kapusta, Maksim Kos, Antonio Lukić, Jerko Šegvić

University of Zagreb, Faculty of Electrical Engineering and Computing  
Unska 3, 10000 Zagreb, Croatia

## Sažetak

Cilj nadopunjavanja slike je nedostajuće piksele slike zamijeniti čim uvjerljivijim sadržajem. U projektu smo opisali i implementirali model za nadopunjavanje slika temeljen na difuziji. Trenirali smo difuzijski model na skupu podataka MNIST, implemntirali RePaint (Lugmayr et al.2022) petlju za popunjavanje slike te razvili sučelje za vizualizaciju maske, nepotpune slike i popunjene slike. Model uvjerljivo ispunjava slike.

## 1. Nadopunjavanje slike

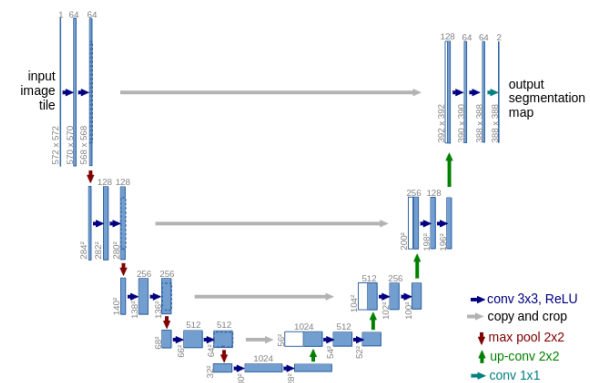
Nadopunjavanje slike (engl. *image inpainting*) je postupak u kojem nadopunjavamo segmente slike koji nedostaju. Nadopuna mora biti semantički smisljena i u skladnom odnosu s poznatim dijelom slike. Zbog ovih čvrstih zahtjeva, modeli za nadopunjavanje slike moraju imati snažne generativne mogućnosti. Generativni difuzijski modeli, uz dobre generativne sposobnosti, omogućuju i manipulaciju latentnog prostora za ugradnju semantičke informacije. Popularni modeli za ovaj zadatak su još i generativne suparničke mreže (engl. *generative adversarial networks*) i autoregresivni modeli. Problem takvih arhitektura je da moraju također naučiti i distribuciju dijelova slike koji nedostaju. U praksi takvi modeli imaju problema sa neobičnim prazninama u slici ili sa prazninama s kojima se nisu susreli prilikom treniranja

## 2. Difuzijski modeli

Difuzijski modeli su su klasa generativnih modela s latentnim varijablama (Ho et al.2020). Ideja difuzijskih modela je simulirati proces reverzne difuzije i naučiti model da iz poznate latentne distribucije konstruira elemente iz distribucije podataka. Svaki difuzijski model se sastoji od dva osnovna elementa: unaprijedni proces i reverzni proces. Unaprijedni proces nema slobodne parametre, već se unaprijed definiranim transformacijama podatak postepeno pretvori u element latentne distribucije. Za distribuciju latentnog prostora odabrali smo normalnu razdiobu. Dimenzija latentnog prostora jednaka je dimenziji prostora podatka. Unaprijedni proces može se promatrati kao markovljev lanac, i moguće je izvesti uvjetne distribuciju sljedećeg stanja na temelju trenutnog. Za distribuciju zadnjeg stanja lanca pretpostavljamo upravo latentnu distribuciju. Unatražnu distribuciju nije traktabilna te nju upravo procjenjujemo pomoću dubokog modela.

### 2.1. UNet

Za predviđanje unatražnog prolaza difuzije koristili smo duboki konvolucijski model arhitekture UNet. Konvolucijski se model trenira tako od šuma dobivenog unaprijednim prolazom difuzije u trenutku  $t$  predviđa ulaznu sliku (trenutak  $t = 0$ ). UNet arhitektura se u području računalnog vida najčešće primjenjuje za semantičku segmentaciju, no može

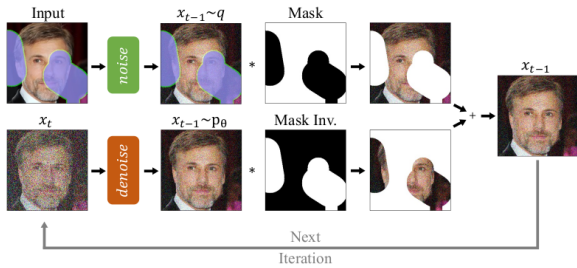


Slika 1: Arhitektura UNet(Ronneberger et al.2015)

se primjenjivati i za druge probleme (npr. difuziju) (Ronneberger et al.2015). Sastoji se od koderskog i dekoderskog dijela, s preskočnim vezama između istih. U koderskom dijelu smanjujemo prostornu rezoluciju, a povećavamo broj kanala dok u dekoderskom dijelu radimo suprotno. Preskočne veze služe kako bi se očuvale prostorno relevantne informacije iz viših slojeva. Koderske i dekoderske blokove čine konvolucijski slojevi i slojevi normalizacije po grupama (engl. batch norm). Konvolucijski blokovi su izvedeni na način da se radi dubinski separabilna konvolucija, koja smanjuje broj parametara i računsku složenost modela. Takvi se slojevi često koriste u modelima namijenjenima za rad na ograničenom sklopovlju (kao što su npr. tableti i pametni mobiteli). Neki od takvih modela su ShuffleNet ((Zhang et al.2017)) i MobileNet ((Howard et al.2017)). Budući da naš model mora moći predvidjeti sliku u svakom koraku difuzije, moramo nekako ugraditi informaciju o vremenskom koraku u model. Kako bi očuvali tu informaciju o vremenskom koraku difuzije  $t$ , koristimo vremenska ugrađivanja. Jednostavnim višeslojnim perceptronom generiramo naučeni vremenski ovisan vektor koji se dodaje latentnoj reprezentaciji.

## 3. RePaint model

RePaint je alogritam za nadpounjavanje slike pomoću pred-treniranog difuzijskog modela. Sam RePaint nema slo-



Slika 2: Slikovni prikaz RePaint algoritma (Lugmayr et al.2022)

bodne parametere te ga zato nije potrebno učiti. To nam omogućava korištenje bilo kojeg difuzijskog modela (Lugmayr et al.2022). Također takav pristup nema pristranost prema nekim oblicima maski. Algoritam funkcionira na sljedeći način. Vidljivi dio slike unaprijednim prolazom preslikamo u stanje  $x_t \sim p_t$ . Nadopunjenu sliku generiramo iz latentnog prostora do stanja  $x_t \sim q_t$ . Sada na nadopunjenu sliku primjenimo inverznu masku i ove dvije slike zbrojimo. Njihov zbroj jednak je novom "latentnom prostoru" iz kojeg generiramo novu nadopunjenu sliku. Ova iteracija se ponavlja dok se cijeli reverzni proces ne izvrši to jest do  $t = 0$

#### 4. Rezultati

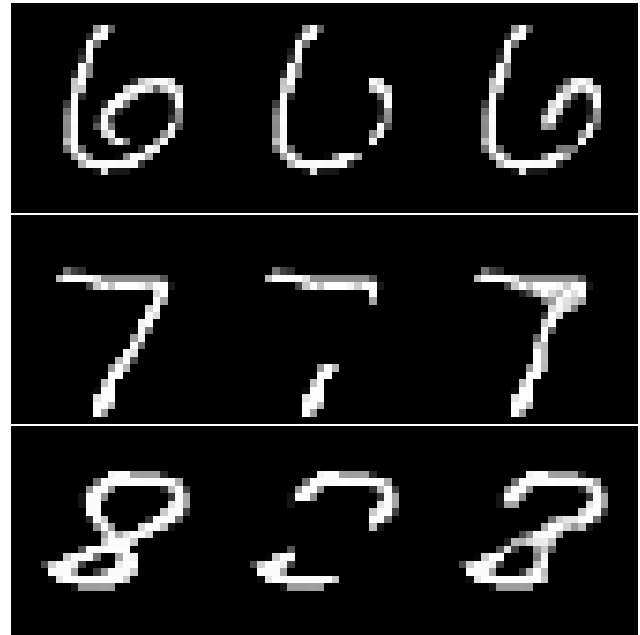
Model smo trenirali na skupu podataka MNIST. Za treniranje smo koristili Nvidia GeForce RTX 3080 grafičku karticu s 10GB memorije. Model smo trenirali kroz 100 epoha s veličinom grupe 64. Optimirali smo kvadratnu razliku između originalne slike i slike dobivenu unatražnim prolazom difuzije. Kako bi testirali funkcionalnost našeg modela, maskirali smo nekoliko uzoraka iz skupa MNIST i nadopunili ih algoritmom RePaint koristeći naučeni model(slika 2). Model je uvjerljivo popunio maskirane djelove slike u većini slučajeva.

#### 5. Zaključak

U ovom radu istražili smo problem nadopunjavanja slike koristeći difuzijske modele. Difuzijski modeli pokazali su se kao obećavajuća klasa generativnih modela zbog svoje sposobnosti učenja latentnih reprezentacija kroz reverzni proces difuzije. Koristeći unaprijedni i reverzni proces, ovi modeli omogućuju postupno dodavanje i uklanjanje šuma na podacima, što omogućuje generiranje realističnih nadopuna slika.

Proveli smo implementaciju modela temeljenog na difuziji za nadopunjavanje slika i testirali ga na skupu podataka MNIST. Eksperimentalni rezultati pokazuju da difuzijski modeli mogu generirati vizualno uvjerljive nadopune čak i za slike s kompleksnim uzorcima nedostajućih područja. Prednosti ovakvog pristupa uključuju:

- **Fleksibilnost maskiranja:** RePaint algoritam, korišten u ovom radu, ne pokazuje pristranost prema određenim oblicima maski, što ga čini generalno primjenjivim na različite vrste ulaznih podataka.



Slika 3: Originalna slika, maskirana slika i rezultati nadopunjavanja

- **Korištenje postojećih modela:** Reverzni proces omogućuje iskorištavanje predtreiniranih difuzijskih modela bez potrebe za dodatnim treniranjem, što značajno smanjuje računalne zahtjeve.
- **Robusnost:** Difuzijski modeli pokazali su robusnost prema nepoznatim obrascima praznina, za razliku od nekih tradicionalnih pristupa poput generativnih suparničkih mreža (GAN-ova).

Međutim, izazovi ostaju u smislu optimizacije performansi i brzine generiranja. Naši eksperimenti sugeriraju potrebu za daljnjim istraživanjem arhitektura difuzijskih modela kako bi se poboljšala kvaliteta nadopuna i smanjilo vrijeme izvođenja reverznih procesa.

Rezultati ovog rada doprinose razumijevanju primjene difuzijskih modela u nadopunjavanju slika i otvaraju mogućnosti za njihovu primjenu u širem spektru zadataka u domeni računalnog vida.

#### Literatura

- Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models.
- Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications.
- Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. 2022. Repaint: Inpainting using denoising diffusion probabilistic models.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation.

Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun.  
2017. Shufflenet: An extremely efficient convolutional  
neural network for mobile devices.