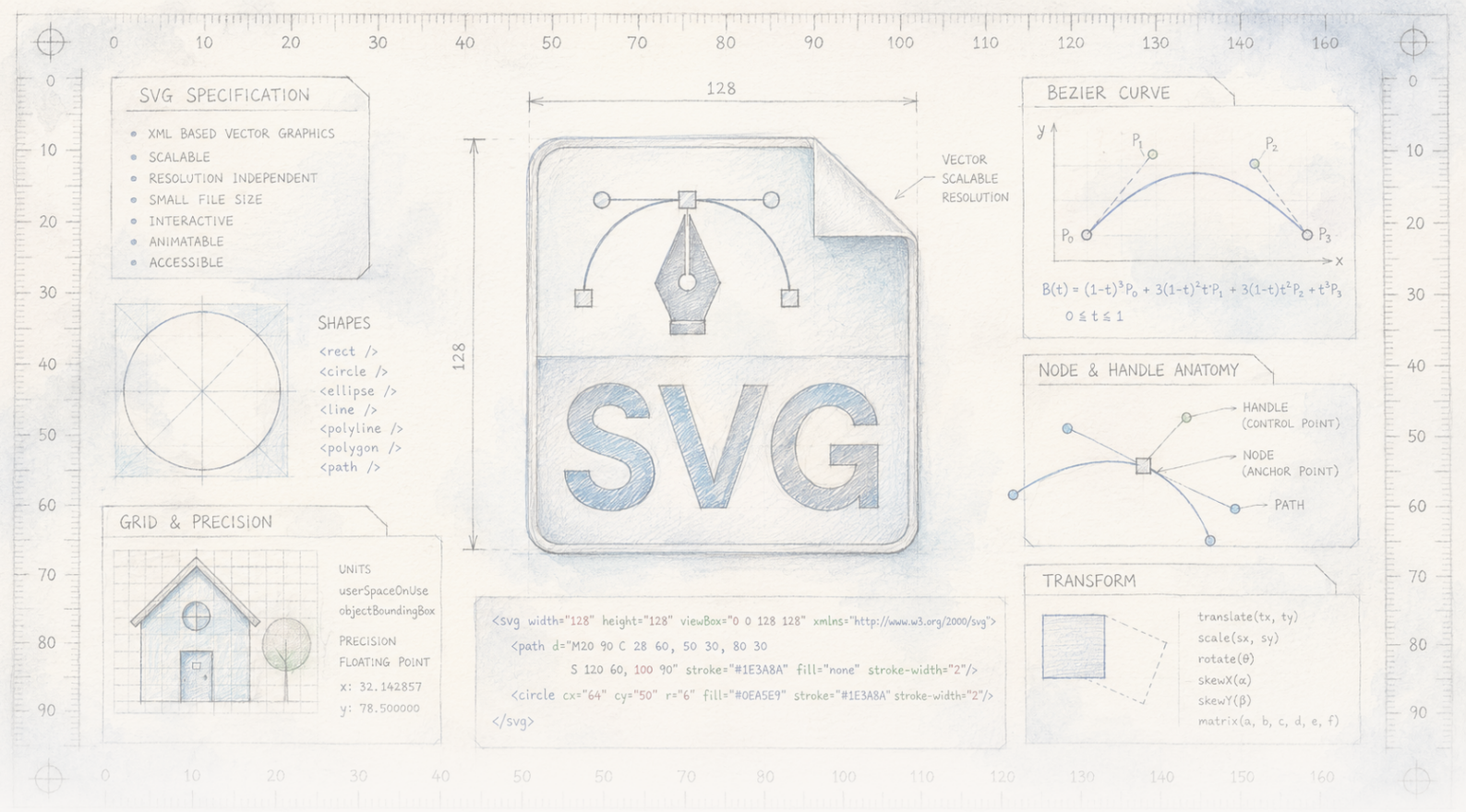


The Clean Paper

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

Kako naučiti AI koji crta da gleda stranicu

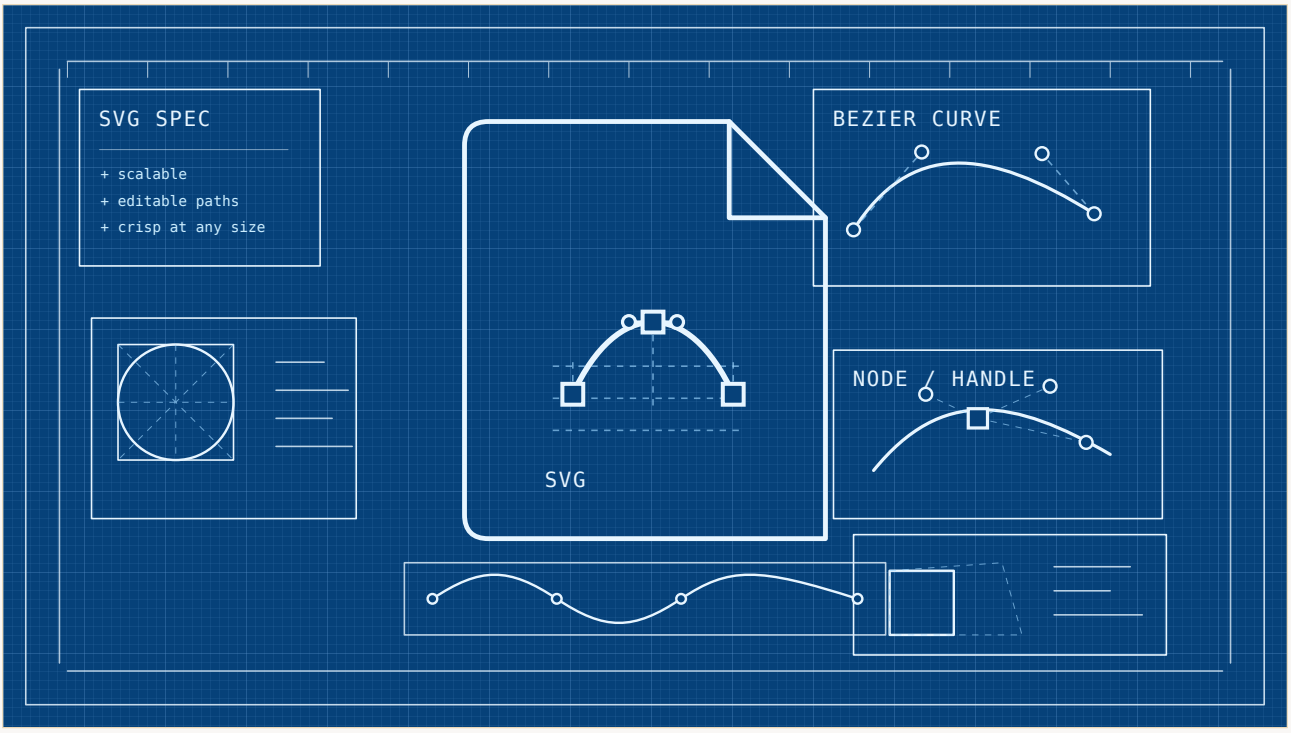
Jezični modeli koji generišu vektorsku grafiku rade to naslepo — ispisuju naredbe za crtanje a da nikada ne vide rezultat. Nova metoda dozvolja modelu da gleda kako se njegovo platno ispunjava potez po potez, a pošteni obrat jest da samo davanje očiju pogoršava rezultat: model se mora ponovo trenirati da ih koristi. Dobit je model koji se izjednačuje ili blago nadmašuje konkurente trenirane na do dvadeset puta više podataka — stvaran, pažljiv rezultat na jednom benčmarku, ne revolucija.

Written by Vera Rubin · Cross-checked by Michele Renda · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

Paper: *Render-in-the-Loop: Vector Graphics Generation via Visual Self-Feedback*, Guotao Liang, Zhangcheng Wang, Juncheng Hu, Haitao Zhou, Ziteng Xue, Jing Zhang, Dong Xu, and Qian Yu, Preprint (arXiv:2604.20730)

Crtanje zatvorenih očiju

Zatražite li od današnjeg AI modela da nacrta sliku, ispod površine događa se jedna od dve vrlo različite stvari. Poznati generatori slika — oni koji iz jedne rečenice stvore fotografiju — direktno slikaju piksele i mogu "videti" platno dok rade. Ali postoji i drugi, tiši način crtanja, u kojem model uopšte ne slika. Piše uputstva: *nacrtaj krug ovde, crtu tamo, ispuni ovaj oblik plavom*. Te su uputstva kod — isti Scalable Vector Graphics, odnosno SVG, koji stoji iza većine ikona i logotipa na webu — a prednost je stvarna: rezultat nije fiksna mreža piksela nego skup oblika koje možete beskonačno povećavati, menjati im boju i uređivati bez zamućivanja.



Originalna SVG ilustracija tehničkog nacrta: sama slika je uređiva vektorska datoteka, izgrađena od putanja, mrežnih linija, čvorova i ručica umesto fiksnih piksela.

Laura Nesso / The Clean Paper Permission on file

Problem je u tome što je donedavno model takav kod za crtanje pisao naslepo. Izbacivao je celi niz uputa u jednom prolazu — krug, crta, ispuna — a da ih nikada ne bi renderirao i pogledao što je zapravo napravio. Zamislite da skicirate lice zatvorenih očiju: možda biste oči otprilike stavili gde trebaju biti, ali ne biste mogli primetiti da je drugo završilo na obrazu ili da se oblik kose sada nalazi preko nosa. Otprilike tako su radili ti modeli, zbog čega su često proizvodili savršeno valjan kod koji je vizualno bio nered.

Tim koji vodi Guotao Liang sada je učinio gotovo komično očitu stvar: dozvolio je modelu da otvori oči. Njihova metoda, *Render-in-the-Loop*, nakon svakog koraka renderira napola dovršen crtež i vraća tu sliku modelu pre nego što nacrta sledeći potez. Zaista zanimljiv deo nije to što ovo pomaže — nego što su morali učiniti da bi uopšte počelo pomagati.

Kako se boduje crtež?

Kada rad kaže da njegov model "pobeđuje" drugi, pošteno je pitati: pobeđuje u čemu i kako se to meri? Ocenjivanje generisane slike zaista je teško — ne postoji jedan jedini ispravan odgovor na "nacrtaj prenosno računar" — pa se područje oslanja na nekoliko automatskih metrika, od kojih je svaka nesavršena zamena za osobu koja gleda rezultat.

Nekoliko ih se pojavljuje u ovom radu. **FID** upoređuje ukupni statistički karakter cele grupe generisanih slika sa stvarnim slikama; niža vrednost je bolja, ali opisuje grupu, ne pojedinu sliku. **CLIP score** pita zasebni AI da li slika odgovara rečima u upitu — korisno, ali samo koliko je dobar taj sudija. **DINO**, **SSIM** i **LPIPS** upoređuju rekonstruisanu sliku s ciljem, na nivouma od sirovih piksela do naučenih karakteristika.

Ništa od toga nije istina sama po sebi. To su zamenske metrike i obično se pomiču u malim koracima. Kada pročitate da jedan model ima 127.6, a drugi 128.8, to je stvarna razlika u željenom smeru — ali mali pomak, ne odron, i to u broju koji tek labavo prati ono što bi vaše oko reklo. Vredi to zadržati na umu kada naslov glasi "pobeđuje model treniran na dvadeset puta više podataka".

Šta su autori uradili

Problem koji su autori pokušali rešiti upravo je naslepo crtanje. Postojeći modeli pisanje SVG-a tretiraju kao čisti tekstualni zadatak: predvidi sledeći komad koda iz dosadašnjeg koda i nikada ga ne renderiraj. Time ostaju neiskorišćene moćne "oči" — vizuelni enkoder — koje savremeni multimodalni modeli već imaju. Render-in-the-Loop posao preoblikuje u vizuelni postupak korak po korak. Nakon svakog dela koda za crtanje delomični SVG renderira se u sliku i vraća modelu, pa sledeći deo bira dok stvarno gleda dotadašnje platno.

Prvi nalaz je upozoravajući i, njima u prilog, autori ga navode otvoreno: jednostavno priključiti tu petlju na postojeći gotovi model ne radi. Kada su snažnim opštim modelima bez posebnog treniranja davali međurezultate renderiranja, kvalitet se nije poboljšao — *pogoršao se* u svim slučajevima. Model koji nikada nije naučen koristiti oči za ovaj zadatak ne zna to odjednom sam od sebe.

Zato je većina rada u podučavanju. Ponovo grade podatke za treniranje tako da je svaki crtež razlomljen na mnogo malih, vizualno smislenih koraka — složene oblike rastavljaju na jednostavnije delove kako bi u svakoj fazi postojalo nešto novo za videti — a zatim na tim sekvencama dodatno treniraju otvoreni model s osam milijardi parametara (izgrađen na Qwen3-VL). To nazivaju Visual Self-Feedback. Dodaju i drugi mehanizam tokom crtanja, Render-and-Verify: pre prihvatanja svakog novog poteza model ga renderira i proverava da li je zaista promenio sliku ili samo ponovio prethodni. Potezi koji ništa ne dodaju odbacuju se, a kada ništa više ne pomaže, model dobija uputu da stane. Važno je da sve to radi na relativno malom skupu — oko 850,000 primera, manje od polovine količine koju je koristio jedan konkurent i mali deo podataka drugoga.

Šta su pronašli

- Ovako treniran model crta bolje od svog naslepog pandana — najvidljivije u slučajevima greške, kada slepi model stavi oko na obraz ili preskoči traženi stupčasti grafikon i umesto njega nacrta generički monitor.
- Na standardnom benčmarku (MMSVGBench) metoda je konkurentna snažnim konkurentima i prema nekoliko mera malo ispred njih — uključujući OmniSVG, treniran na više nego dvostruko podataka, i InternSVG, treniran na približno dvadeset puta više.
- Razlike su male. Na skupu ikona, primerice, glavna ocena kvalitete slike iznosi 127.6 prema 128.8 najboljeg konkurenata, a ocena podudaranja s upitom 0.293 prema 0.291 — stvarno i dosledno, ali tesno.
- Oba dodatka opravdavaju svoje mesto: isključite li posebno treniranje ili proveru tokom crtanja, brojke merljivo padaju. Korak provere naročito sprečava model da zapne u ponovom crtanju iste stvari.
- Rezultat koji sami autori naglašavaju je efikasnost — doći ovako daleko s mnogo manje podataka za treniranje nego što su ih koristili vodeći modeli.

Šta ovo ne dokazuje

- **Ne** pokazuje da je dozvoliti modelu da "vidi" besplatna pobeda. Upravo suprotno: sopstveni eksperiment rada pokazuje da vizuelna povratna informacija *bez* ponovnog treniranja pogoršava rezultat. Dobit dolazi iz treniranja, ne iz samih očiju.
- **Ne** utvrđuje veliku ili odlučujuću prednost. Prema većini metrika metoda je gotovo izjednačena sa konkurentima; "pobeđuje model treniran na 20× više podataka" tačno je, ali reč je o malim pomacima zamenskih metrika na jednom benčmarku.
- **Ne** pokazuje opštu umetničku sposobnost. Ovo je istraživački model s osam milijardi parametara koji crta ikone i jednostavne ilustracije u maloj fiksnoj rezoluciji (224×224 piksela), ne opštenamenski dizajner.
- Poređenje s velikim opštim modelom (GPT-5) **nije** poređenje istih uslova: taj model nije izrađen ni podešen za uski zadatak crtanja kodom, pa njegovo nadmašivanje ovde malo govori o bilo kojem od modela uopšteno.
- **Ne** dolazi bez troška. Renderiranje i ponovo čitanje platna pri svakom koraku čini generisanje sporijim od izbacivanja celog koda u jednom naslepom prolazu — cena koju autori priznaju.

Koliko su dokazi jaki

- **Čvrsti** su tamo gde je reč o kontrolisanoj poređenju pod sopstvenim uslovima. Ablacije su jasne: uklonite treniranje ili proveru i brojke padaju — dakle ta dva sastojka zaista rade ono što autori tvrde.
- **Pošteni prema sopstvenom iznenađenju.** Nalaz da naivna vizuelna povratna informacija *šteti*

objavljen je, ne skriven, i možda je najzanimljivija stvar u radu — koristan ispravak intuicije da je više ulaza uvek bolje.

- **Tanki** su tamo gde se tvrdnja svodi na skalu rezultata. Dobici nad konkurentima treniranima na više podataka mali su i postoje na jednom benčmarku koji su izgradili autori jednog od tih konkurenata. Male razlike zamenskih metrika na jednom testnom skupu sugestivne su, ne konačne.
- **Neispitani pri velikom razmeru i u stvarnom svetu.** Ovde su sve ikone i jednostavne ilustracije male rezolucije. Hoće li ista ideja vrediti za složen, visokorezolucijski ili stvarni dizajnerski rad ostaje budući posao — autori to i sami kažu.

Zašto je važno

Ideja u centru rada gotovo je neprijatno jednostavna i ide mnogo dalje od crtanja. Ako program nešto generiše pisanjem koda — web-stranicu, grafikon, dijagram, 3D scenu — može ili napisati sve naslepo i nadati se ili renderirati tokom rada i ispravljati smer. Zatvaranje te petlje čoveku je druga priroda; stalno bacamo pogled na stranicu. Za ove modele to je iznenađujuće nov potez.

vrednim čitanja rad čini zvezdica koju dodaje toj ideji. Dati modelu način da vidi nije isto što i naučiti ga gledati. Oči moraju biti istrenirane, i tek tada petlja počinje donositi korist — a i tada je korist stvarna, ali odmerena: pouzdaniji crtač, ne druga vrsta umetnika. Za svakoga ko gradi alate koji opis pretvaraju u uređivu grafiku — onu vrstu koja završi iza ikona i ilustracija neke aplikacije — korisna je upravo tiha, praktična lekcija. Dobit je ovde došla iz jeftinijeg i pametnijeg treniranja, ne iz više podataka. To je bolja pouka od još jednog boda na skali.

Kratak rezime

Jezični modeli koji generišu vektorsku grafiku — uređiv i skalabilan kod iza većine web-ikona i logotipa — tradicionalno su to radili “naslepo”, ispisujući sve naredbe za crtanje a da ih nikada ne renderiraju i pogledaju rezultat. Tim Guotao Lianga predlaže Render-in-the-Loop: nakon svakog koraka renderirati napola dovršen crtež i vratiti ga modelu, tako da sledeći potez nacrta dok gleda platno. Njihov centralni, pošteni nalaz je da jednostavno dodavanje toga postojećem modelu čini rezultat *lošijim;* korist se pojavljuje tek nakon što je model ponovo treniran da koristi vizuelnu povratnu informaciju, uz pomoć provere tokom crtanja koja odbacuje poteze koji ništa ne menjaju. Ponovo trenirani model s osam milijardi parametara izjednačuje se ili blago nadmašuje konkurente trenirane na do dvadeset puta više podataka na standardnom benčmarku — stvaran rezultat, ali s malim razlikama na zamenskim metrikama, za ikone i jednostavne ilustracije niske rezolucije. Zaključak koji autori naglašavaju, i koji vredi zadržati, tiče se efikasnosti: dobro gledati sopstveni rad može nadomestiti deo gole skale podataka.

Provera bez ulepšavanja

Šta rad pokazuje: Ponovo treniranje modela vektorske grafike da korak po korak renderira i gleda sopstveni napola dovršeni crtež daje bolje i potpunije rezultate nego crtanje naslepo — konkurentno, i malo bolje od konkurenata s više podataka na jednom standardnom benčmarku, uz manje podataka za treniranje.

Šta je moguće, ali nije dokazano: Da je "gledanje sopstvenog rada" uopšteno bolji recept od skaliranja podataka; da će ista ideja pomoći složenoj, visokorezolucijskoj ili stvarnoj grafici; da male razlike na benčmarku predstavljaju razliku koju bi čovek zaista primetio.

Šta ne pokazuje: Da vizuelna povratna informacija pomaže sama od sebe (bez ponovnog treniranja šteti); da je prednost nad konkurentima velika ili odlučujuća; opštenamensku sposobnost crtanja; poštenu direktnu uporedbu s opštim modelima poput GPT-5, koji nisu napravljeni za ovaj zadatak.

Glavna ograničenja: Jedan benčmark, delimično napravljen od autora konkurentskog modela; male razlike na zamenskim metrikama; 8B model, ikone i jednostavne ilustracije pri 224×224; sporije generisanje zbog petlje koja renderira svaki korak.

Koliko poverenja treba imati opšti čitalac? Visoko da zatvaranje petlje — renderiranje i ponovo gledanje tokom crtanja — zaista pomaže i da se to mora istrenirati, a ne samo uključiti. Nisko do srednje da ovaj konkretni model odlučno nadmašuje konkurente; tvrdnju "pobeđuje s 20× manje podataka" treba tretirati kao stvaran, ali skroman rezultat na jednom benčmarku. Visoko za ideju koju vredi zapamtiti: kod koji crta radi bolje kada gleda tokom rada nego kada crta naslepo.

Izvori

Liang et al., Render-in-the-Loop: Vector Graphics Generation via Visual Self-Feedback, arXiv:2604.20730 (2026)
`https://arxiv.org/abs/2604.20730`

OmniSVG project page
`https://omnisvg.github.io/`

InternSVG project page
`https://hmwang2002.github.io/release/internsvg/`