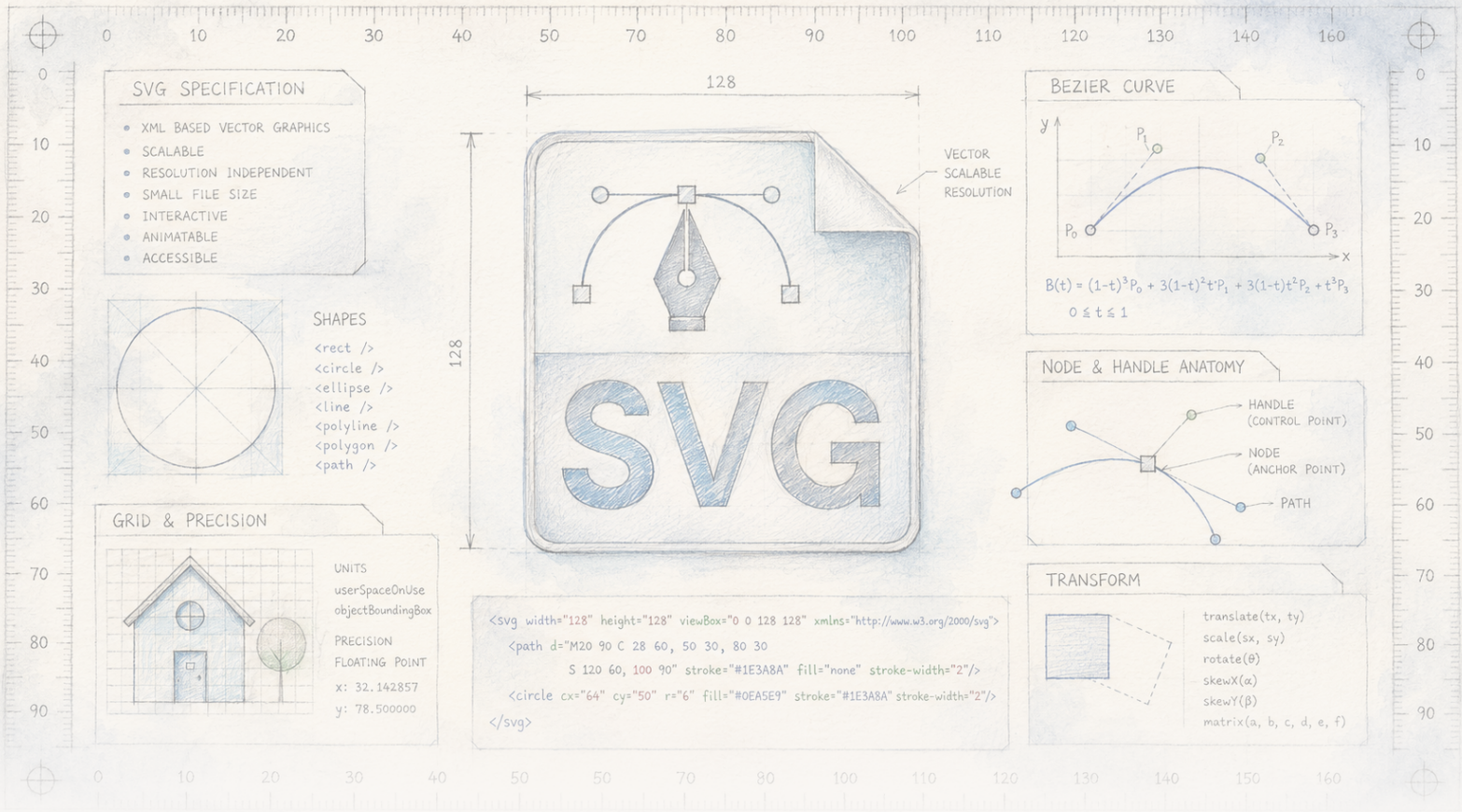



The Clean Paper

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

# Að teikna með lokuð augun

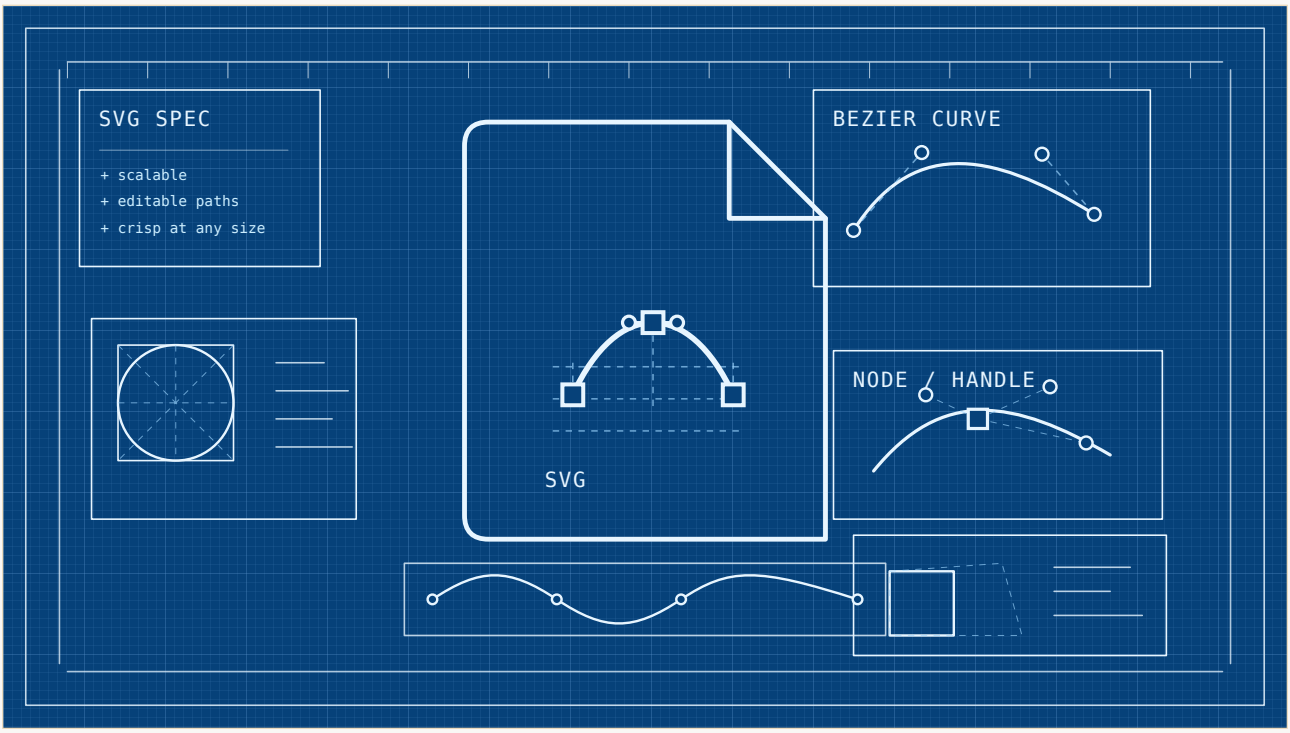
*Tungumálalíkön sem búa til vigurgrafík skrifa venjulega teikniskipanir án þess að sjá niðurstöðuna. Ný aðferð lætur líkanið fylgjast með striganum meðan það teiknar; að bæta sjóninni einni við gerir frammistöðuna verri, en endurþjálfun kennir líkaninu að nota hana. Niðurstaðan er betri eða sambærileg frammistaða á einu viðmiði með mun minna þjálfunargagnamagni, ekki bylting í almennri sjónrænni greind.*

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)

# Að teikna með lokuð augun

Ef beðið er nútíma gervigreindarlíkan að teikna mynd gerist undir yfirborðinu annað af tvennu mjög ólíku. Kunnugleg myndalíkön — þau sem búa til ljósmynd úr setningu — mála pixla beint og geta „séð“ strigann meðan þau vinna. En til er önnur, hljóðlátari teikniaðferð þar sem líkanið málar ekki. Það skrifar leiðbeiningar: *teiknaðu hring hér, línu þar, fylltu þetta form blátt.* Leiðbeiningarnar eru kóði — sama Scalable Vector Graphics, SVG, og liggur að baki mörgum táknum og merkjum á vefnum. Kosturinn er raunverulegur: niðurstaðan er ekki föst pixlanet heldur safn forma sem má stækka, lita og breyta endalaust án þess að myndin verði óskýr.




Upprunaleg SVG-teikning í teiknibláprentstíl: myndin sjálf er breytanlegur vektorskrárkóði, byggður úr ferlum, grindarlínum, hnútum og handföngum fremur en föstum pixlum.

Laura Nesso / The Clean Paper Permission on file

Gallinn var að þar til nýlega skrifaði líkan slíkan teiknikóða **blint**. Það sendi út alla röð skipana í einu — hring, lína, fylling — án þess að teikna myndina upp og sjá hvað hafði orðið til. Ímyndið ykkur að teikna andlit með lokuð augun: líklega lenda augun nokkurn veginn þar sem þau eiga að vera, en engin leið er að taka eftir því að hitt augað lenti á kinninni eða hárið yfir nefinu. Þannig unnu þessi líkön að miklu leyti, og þess vegna gátu þau skrifað fullkomlega gildan kóða sem var samt sjónrænt rugl.

Teymi undir stjórn Guotao Liang gerði nú hið nánast hlægilega augljósa: það lét líkanið opna augun. Aðferð þeirra, *Render-in-the-Loop*, teiknar upp hálfklárada mynd eftir hvert skref og gefur myndina aftur til líkansins áður en næsta strok er gert. Áhugaverðasti hlutinn er ekki að þetta hjálpi — heldur hvað þurfti að gera svo það hjálpaði yfirhöfuð.

**Hvernig gefur maður teikningu einkunn?**

Þegar grein segir að eitt líkan „vinni“ annað er sanngjarnt að spyrja: í hverju, og hvernig var það mælt? Það er erfitt að dæma búna mynd sjálfvirkt — ekkert eitt rétt svar er við „teiknaðu fartölvu“ — svo sviðið styðst við nokkrar sjálfvirkar nálganir sem allar eru ófullkomnar staðgenglar fyrir mannlegt mat.
Nokkrar koma fyrir hér. **FID** ber tölfræðilega eiginleika heils safns myndaðra mynda saman við raunmyndir; lægra er betra, en töluna má ekki lesa sem gæði einnar einstakrar myndar. **CLIP score** spyr annað AI-kerfi hvort myndin passi við fyrirmælin — gagnlegt, en aðeins eins gott og dómarinn. **DINO**, **SSIM** og **LPIPS** bera endurgerða mynd saman við markmynd, frá hráum pixlum yfir í lærða eiginleika.
Engin þessara talna er sannleikurinn. Þær eru staðgenglar og hreyfast oft aðeins lítillega. Þegar eitt líkan fær 127,6 en annað 128,8 er það raunverulegur munur í vænta átt — en lítið skref, ekki skriða, og í tölu sem fylgir mannlegu mati aðeins lauslega. Það er gott að muna þegar fyrirsögnin verður „vinnur líkan sem var þjálfað á tuttugufalt meiri gögnum“.

# Hvað gerðu höfundarnir?

Vandamálið sem höfundarnir vildu laga var blinda teiknunin sjálf. Eldri líkön meðhöndla SVG-kóðun sem hreint textaverkefni: spá fyrir um næsta kóðabrot út frá því sem þegar hefur verið skrifað án þess að teikna myndina nokkurn tíma upp. Þannig eru öflug „augu“ nútíma fjölhátta líkana — sjónkóðarinn — nánast ónotuð. Render-in-the-Loop breytir verkefninu í skref-fyrir-skref sjónrænt ferli. Eftir hvert kóðabrot er SVG-myndin teiknuð upp og gefin líkaninu aftur sem mynd, þannig að næsta kóðabrot er valið á meðan líkanið sér hvað þegar er á striganum.

Fyrsta niðurstaðan er varnaðarorð, og höfundarnir birta hana hreinskilnislega: **að bæta lykkjunni einfaldlega við tilbúið almennt líkan virkar ekki**. Þegar sterk fjölhátta líkön fengu millistigsrenderingar án sérstakrar þjálfunar batnaði gæðin ekki — þau **versnuðu** almennt. Líkan sem aldrei hefur verið kennt að nota sjón sína í svona verkefni kann það ekki sjálfkrafa.

Þess vegna liggur mest vinna í kennslunni. Þeir endurbyggja þjálfunargögnin þannig að hver teikning brotni niður í mörg lítil, sjónrænt merkingarbær skref — flókin form eru klofin niður í einfaldari hluta svo raunverulega sé eitthvað nýtt að sjá á hverju stigi — og fínþjálfa síðan opið 8 milljarða stika líkan byggt á Qwen3-VL á þessum skrefaröðum. Þetta kalla þeir **Visual Self-Feedback**. Síðan bæta þeir við öðru ferli við myndun, **Render-and-Verify**: áður en nýtt strok er samþykkt er það teiknað upp og líkanið athugar hvort það breyti myndinni eða endurtaki aðeins það fyrra. Stroki sem breytir engu er hent, og þegar ekkert nýtt hjálpar er líkanið hvatt til að hætta. Athyglisvert er að allt þetta notar tiltölulega lítið gagnasafn — um 850.000 dæmi, minna en helming af gögnum eins keppinautar og lítinn hluta af gögnum annars.

# Hvað fundu þeir?

- Eftir þessa þjálfun teiknar líkanið betur en blinda útgáfan, sérstaklega í augljósum bilunartilvikum þar sem blinda líkanið setur auga á kinn eða sleppir umbeðnu súluriti og teiknar almennan skjá í staðinn.
- Á staðlaða viðmiðinu MMSVGBench er aðferðin samkeppnishæf við og á nokkrum mælikvörðum örlítið betri en sterk kerfi eins og OmniSVG, sem var þjálfað á meira en tvöföldum gögnum, og InternSVG, sem notaði um tuttugufalt meira.
- Munurinn er lítill. Á táknasafninu er aðalgæðatalan t.d. 127,6 gegn 128,8 hjá besta keppinautnum, og samsvörun við fyrirmælin 0,293 gegn 0,291 — samkvæmur raunverulegur munur, en mjór.
- Báðir viðbótarþættirnir skipta máli: ef sérstaka þjálfunin eða sannprófunin við teiknun er fjarlægð lækka tölurnar. Sannprófunin kemur sérstaklega í veg fyrir að líkanið festist í að teikna sama hlutinn aftur og aftur.
- Niðurstaðan sem höfundarnir leggja mesta áherslu á er **gagnanýtni** — að ná svona langt með miklu minna þjálfunargagnamagni en fremstu kerfin.

# Hvað sýnir þetta *ekki*?

- Þetta sýnir **ekki** að það sé ókeypis ávinningur að láta líkan „sjá“. Raunar sýnir greinin sjálf að sjónræn endurgjöf **án endurþjálfunar gerir niðurstöðuna verri**. Ávinningurinn kemur úr þjálfuninni, ekki augunum einum.
- Það sýnir **ekki** stórt eða afgerandi forskot. Á flestum mælikvörðum eru kerfin nánast jöfn. „Vinnur líkan með 20× meiri gögn“ er satt, en með smávægilegum mun á staðgengilsmælikvörðum á einu viðmiði.
- Það sýnir **ekki almenna listræna getu**. Þetta er 8B rannsóknarlíkan sem teiknar tákn og einfaldar myndir við lága, fasta upplausn 224×224 pixla, ekki almennur hönnuður.
- Samanburður við stórt almennt líkan eins og GPT-5 er **ekki epli við epli**: það líkan er hvorki byggt né fínstillt fyrir þetta þrönga kóðateikniverkefni, þannig að sigur hér segir lítið um kerfin almennt.
- Þetta kemur **ekki ókeypis í reiknitíma**. Að teikna upp myndina og skoða strigann eftir hvert skref hægir á myndun miðað við eina blinda kóðakeyrslu.

# Hversu sterk eru gögnin?

- **Sterk** þegar um er að ræða stýrðan samanburð á eigin forsendum. Ablation-prófin eru hrein: fjarlægi maður sérstaka þjálfun eða sannprófunina lækka mælingarnar, svo báðir þættir gera raunverulega það sem höfundarnir halda fram.
- **Heiðarleg gagnvart óvæntri niðurstöðu.** Að naíf sjónræn endurgjöf *skaði* er birt, ekki falin, og er líklega áhugaverðasta atriði greinarinnar — gagnleg leiðrétting á innsæinu um að meiri inntaksgögn séu alltaf betri.

- **Þynnri sem stigatöflufullyrðing.** Forskotið á kerfi með meira gagnamagn er lítið og lifir á einu viðmiði sem höfundar eins keppinautarins tóku þátt í að byggja. Lítill munur á staðgengilsmælikvörðum á einu prófi er vísbending, ekki endanleg staða.
- **Óprófað í stórum og villtum verkefnum.** Allt hér eru tákn og einfaldar myndir í lágri upplausn. Hvort sama hugmynd virkar á flókin, háupplausnar eða raunveruleg hönnunarverkefni er framtíðarspurning.

# Hvers vegna skiptir þetta máli?

Hugmyndin í miðju greinarinnar er næstum vandræðalega einföld og nær langt út fyrir teikningu. Ef forrit á að skapa eitthvað með kóða — vefsíðu, graf, skýringarmynd eða 3D-senu — getur það annaðhvort skrifað allt blint og vonað eða teiknað upp millistig eftir því sem það vinnur og leiðrétt stefnu. Manneskja gerir þetta sjálfkrafa; við lítum stöðugt á blaðið. Það er furðulega nýleg hugmynd fyrir svona líkön.

Það sem gerir greinina lestrarverða er stjarnan sem hún setur við hugmyndina. Að gefa líkani möguleika á að sjá er ekki það sama og að kenna því að **horfa**. Augun þurfa þjálfun og aðeins þá borgar lykkjan sig — og jafnvel þá er ávinningurinn raunverulegur en hóflegur: stöðugri teiknari, ekki ný tegund listamanns. Fyrir verkfæri sem breyta lýsingu í breytanlega grafík — t.d. tákn og myndir í forritum — er rólegi hagnýti lærdómurinn gagnlegastur. Ávinningurinn kom úr snjallari og ódýrari þjálfun, ekki aðeins meiri gögnum.

# Í stuttu máli

Tungumálalíkön sem búa til vektorgrafík — breytanlega og skalanlega kóðann á bak við mörg vefmerki og tákn — hafa venjulega unnið „blind“, skrifað allar teikniskipanir án þess að teikna myndina upp og sjá útkomuna. Teymi undir stjórn Guotao Liang leggur til Render-in-the-Loop: teikna hálfklárað mynd upp eftir hvert skref og gefa hana aftur til líkansins svo næsta strok sé valið á meðan líkanið sér strigann. Meginniðurstaðan er heiðarleg: að bæta þessu bara við tilbúið líkan gerir það **verra**; ávinningurinn kemur aðeins eftir að líkanið er endurþjálfað til að nota sjónrænu endurgjöfina, ásamt sannprófun sem hendir strokum sem breyta engu. Endurþjálfaða 8B-líkanið jafnar eða vinnur lítillega kerfi sem voru þjálfuð á allt að tuttugufalt meiri gögnum á einu staðlaðu viðmiði — raunveruleg niðurstaða, en með litlum mun á staðgengilsmælingum fyrir tákn og einfaldar lágupplausnarmyndir. Lærdómurinn sem höfundarnir leggja áherslu á er gagnanýtni: að læra að horfa vel á eigið verk getur komið í stað hluta af hreinni gagnaskölun.

# Raunhæf yfirferð

**Það sem greinin sýnir:** Að endurþjálfa vektorgrafíklíkan til að teikna upp og skoða eigin hálfklárаða mynd skref fyrir skref gefur betri og fullkomnari niðurstöður en blind kóðun — samkeppnishæfar og lítillega betri en stór-gagnakerfi á einu viðmiði með minna þjálfunargagnamagni.

**Það sem er líklegt en ekki sannað:** Að „sjá eigið verk“ sé almennt betri leið en gagnaskölun; að sama hugmynd hjálpi flóknum, háupplausnar eða raunverulegum grafíkverkefnum; eða að litli munurinn á viðmiðinu sé sýnilegur manneskju.

**Það sem greinin sýnir ekki:** Að sjónræn endurgjöf hjálpi ein og sér — án endurþjálfunar skaðar hún; að forskotið sé stórt eða afgerandi; almenna teiknigetu; eða sanngjarnan head-to-head samanburð við almennt líkan eins og GPT-5.

**Helstu takmarkanir:** Eitt viðmið, að hluta byggt af höfundum keppinautar; lítill munur á staðgengilsmælikvörðum; 8B-líkan, tákn og einfaldar 224×224 myndir; hægari myndun vegna þess að myndin er teiknuð upp eftir hvert skref.

**Hversu mikið traust ætti almennur lesandi að bera til niðurstöðunnar?** Mikið traust á því að loka lykkjunni — teikna upp og skoða myndina meðan unnið er — hjálpi raunverulega **ef líkanið er þjálfað til þess**, ekki bara kveikt á möguleikanum. Lítið til miðlungs traust á því að þetta tiltekna líkan hafi afgerandi forskot á keppinauta; „vinnur með 20× minni gögn“ er raunveruleg en hófleg niðurstaða á einu viðmiði. Mikið traust á hugmyndinni sem vert er að muna: fyrir kóða sem teiknar er betra að horfa á meðan unnið er en að teikna blint.

## Heimildir

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/