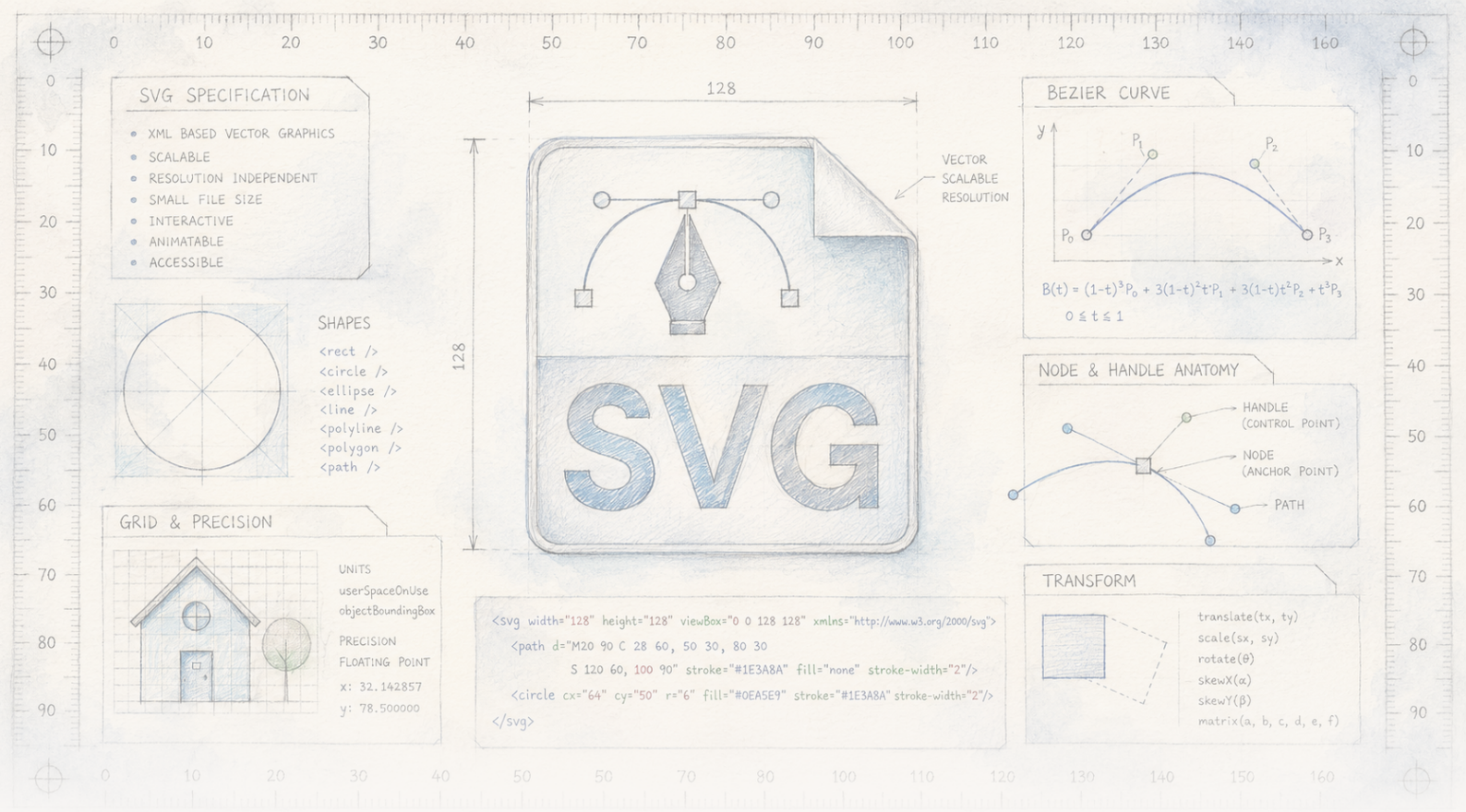


The Clean Paper

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

Rəsm çəkən AI-a səhifəyə baxmağı öyrətmək

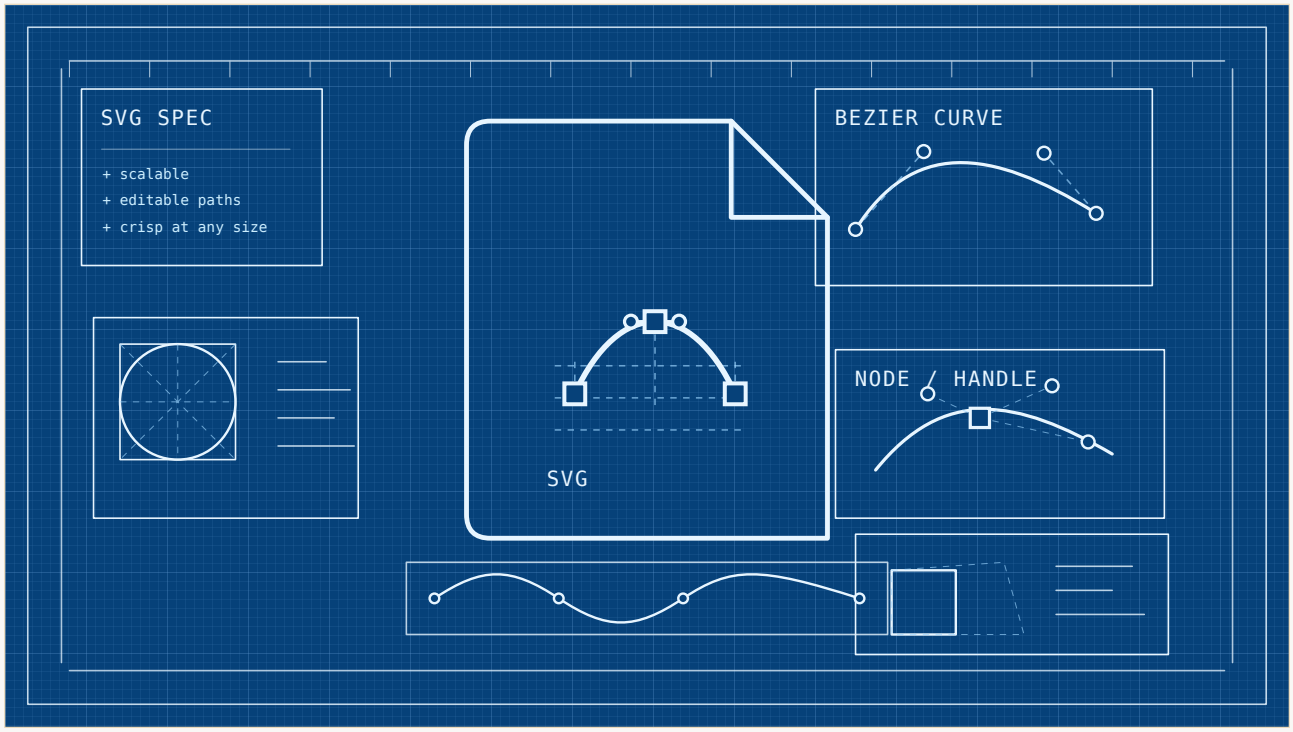
Vektor qrafikası yaradan dil modelləri bunu kor-koranə edir — rəsm əmrlərini nəticəni heç görmədən yazırlar. Yeni metod modelə öz kətanının xətt-xətt dolmasını izləməyə imkan verir və dürüst dönüş budur ki, sadəcə ona göz vermək nəticəni pisləşdirir: bu gözlərdən istifadə etməyi yenidən öyrətmək lazımdır. Nəticə iyirmi dəfəyə qədər çox məlumatla öyrədilmiş rəqiblərə çatan və ya az fərqlə onları keçən modeldir — bir benchmarkda real, diqqətli nəticədir, inqilab deyil.

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)

Gözləriniz bağlı halda rəsm çəkmək

Bugünkü AI modellərindən birindən şəkil çəkməsini istəyin və pərdə arxasında iki çox fərqli şeydən biri baş verir. Tanış şəkil generatorları — bir cümlədən foto yaradanlar — pikselləri birbaşa "boyayır" və işlədikcə kətanı görə bilir. Amma ikinci, daha sakit rəsm növü var: model ümumiyyətlə boyamır. Təlimatlar yazır: *burada dairə çək, orada xətt çək, bu formanı mavi ilə doldur*. Bu təlimatlar koddur — internetdəki əksər ikon və loqoların arxasında duran Scalable Vector Graphics, yəni SVG — və bunun real üstünlüyü var: nəticə sabit piksel toru deyil, bulanıqlaşmadan istədiyiniz qədər böyüdə, rəngini dəyişə və redaktə edə biləcəyiniz formalar toplusudur.



Orijinal SVG blueprint təsviri: şəkilin özü sabit piksellərdən yox, path-lərdən, tor xətlərindən, düyünlərdən və idarəetmə tutacaqlarından qurulan redaktə edilə bilən vektor faylıdır.

Laura Nesso / The Clean Paper Permission on file

Problem ondadır ki, yaxın vaxtlara qədər belə rəsm kodunu yazan model bunu kor-koranə edirdi. Təlimatların bütün ardıcıllığını bir dəfəyə — dairə, xətt, doldurma — çıxarır, nə yaratdığını görmək üçün heç vaxt render etmirdi. Gözləriniz bağlı üz cızdığınızı təsəvvür edin: gözləri təxminən lazımi yerə qoya bilərsiniz, amma ikincisinin yanağa düşdüyünü və ya saç üçün çəkdiyiniz formanın burnun üstünü örtdüyünü görə bilməzsiniz. Bu modellər təxminən belə işləyirdi; buna görə də sintaktik baxımdan tam düzgün, amma vizual olaraq qarışıq kod yaratmaları tez-tez baş verirdi.

Guotao Liang-ın rəhbərlik etdiyi komanda indi demək olar komik dərəcədə aşkar addımı atıb: modelə gözlərini açmağa imkan verib. Onların *Render-in-the-Loop* metodu hər addımdan sonra yarımçıq rəsmi render edir və növbəti xətti çəkməzdən əvvəl həmin şəkli modelə geri

verir. Həqiqətən maraqlı hissə bunun kömək etməsi deyil — ümumiyyətlə kömək etməsi üçün nə etməli olduqlarıdır.

Rəsmə necə bal verilir?

Məqalə bir modelin digərini “məğlub etdiyini” deyəndə haqlı olaraq soruşmaq olar: nədə və necə ölçülüb? Generasiya olunmuş şəkli qiymətləndirmək həqiqətən çətindir — “noutbuk çək” tapşırığının yeganə doğru cavabı yoxdur — buna görə sahə bir neçə avtomatik ölçüyə güvənir; onların hər biri insanın nəticəyə baxmasının qüsurlu əvəzidir.

Bu məqalədə bir neçəsi görünür. **FID** yaradılmış şəkillərin bütöv dəstinin statistik görünüşünü real şəkillərlə müqayisə edir; aşağı daha yaxşıdır, amma ayrı şəkli yox, dəsti təsvir edir. **CLIP score** ayrıca AI-dan şəklin promptdakı sözlərə uyğun olub-olmadığını soruşur — faydalıdır, amma həmin hakimin özü qədər dəqiqdir. **DINO**, **SSIM** və **LPIPS** bərpa olunmuş şəkli hədəflə xam piksellərdən öyrənilmiş xüsusiyyətlərə qədər müxtəlif səviyyələrdə müqayisə edir.

Bunların heç biri həqiqət deyil. Onlar dolayı göstəricilərdir və çox vaxt kiçik addımlarla dəyişir. Bir modelin 128.8 yerinə 127.6 bal aldığını oxuyanda bu, nəzərdə tutulan istiqamətdə real fərqdir — amma uçurum deyil, yalnız insan gözünün deyəcəyini təxmini izləyən bir rəqəmdə kiçik üstünlükdür. Başlıq “iyirmi dəfə çox məlumatla öyrədilmiş modeli məğlub edir” olanda bunu yadda saxlamağa dəyər.

Müəlliflər nə etdilər

Müəlliflərin düzəltmək istədiyi problem məhz kor rəsm idi. Mövcud modellər SVG yazmağı sırf mətn tapşırığı kimi qəbul edir: indiyə qədərki koddan növbəti kod parçasını proqnozlaşdır və heç vaxt render etmə. Beləliklə, müasir multimodal modellərin onsuz da daşıdığı güclü “gözlər” — görmə enkoderi — boş qalır. Render-in-the-Loop işi addım-addım vizual proses kimi yenidən qurur. Rəsm kodunun hər parçasından sonra qismən SVG şəkilə render olunur və modelə təsvir kimi geri verilir; beləliklə, növbəti parça indiyə qədərki kətana həqiqətən baxaraq seçilir.

Onların ilk nəticəsi xəbərdarlıqdır və bunu açıq demələri müəlliflərin xeyrinədir: sadəcə bu dövrəni hazır modelə əlavə etmək işləmir. Xüsusi təlim olmadan güclü ümumi modellərə aralıq renderlər verəndə keyfiyyət yaxşılaşmadı — bütün hallarda *pisləşdi*. Bu tapşırıqda gözlərindən istifadə etməyi heç vaxt öyrənməmiş model birdən-birə bunu bacarmır.

Buna görə işin böyük hissəsi öyrətməkdədir. Onlar təlim məlumatlarını elə yenidən qururlar ki, hər rəsm çoxsaylı kiçik, vizual mənalı addımlara bölünsün — mürəkkəb formalar daha sadə hissələrə ayrılır ki, hər mərhələdə görməyə yeni nəsə olsun — və sonra səkkiz milyard parametrli açıq modeli (Qwen3-VL əsasında) bu addım-addım ardıcıllıqlarda fine-tune edirlər. Buna Visual Self-Feedback deyirlər. Rəsm zamanı ikinci mexanizm də əlavə olunur: Render-and-Verify. Hər yeni xətti qəbul etməzdən əvvəl model onu render edib şəkli həqiqətən

dəyişdirib-dəyişdirmədiyini, yoxsa əvvəlkini sadəcə təkrar edib-etmədiyini yoxlayır. Heç nə əlavə etməyən xətlər atılır; daha heç nə kömək etməyəndə modelə dayanması deyilir. Maraqlıdır ki, bütün bunlar kifayət qədər kiçik məlumat dəstində işləyir — təxminən 850,000 nümunə; rəqiblərdən birinin istifadə etdiyinin yarısından azı və digərinin kiçik bir hissəsi.

Nə tapdılar

- Bu şəkildə öyrədilən model kor variantından daha yaxşı çəkir — xüsusən uğursuz nümunələrdə fərq görünür: kor model gözü yanağa qoyanda və ya istənən sütun qrafikini atlayıb əvəzində generic monitor verəndə.
- Standart benchmarkda (MMSVGBench) metod güclü rəqiblərlə rəqabətlidir və bir neçə ölçüdə az fərqlə öndədir — o cümlədən iki dəfədən çox məlumatla öyrədilmiş OmniSVG və təxminən iyirmi dəfə çox məlumatla öyrədilmiş InternSVG.
- Fərqlər kiçikdir. Məsələn, ikon dəstində əsas şəkil-keyfiyyəti balı ən yaxşı rəqibin 128.8-nə qarşı 127.6, prompt uyğunluğu isə 0.291-ə qarşı 0.293-dür — real və ardıcıl üstünlük, amma cüzi.
- Hər iki əlavə hissənin faydası var: xüsusi təlimi və ya rəsm-zamanı yoxlamanı söndürəndə göstəricilər ölçülə bilən dərəcədə düşür. Xüsusilə yoxlama addımı modelin eyni şeyi təkrar-təkrar çəkmə dövrəsinə düşməsinin qarşısını alır.
- Müəlliflərin özlərinin vurğuladığı nəticə səmərəlilikdir — liderlərin istifadə etdiyi təlim məlumatından xeyli azla bu səviyyəyə çatmaq.

Bu nəyi sübut etmir

- Modelə “görmək” imkanı verməyin **pulsuz üstünlük** olduğunu göstərmir. Əksinə: məqalənin öz təcrübəsi göstərir ki, yenidən təlim olmadan vizual feedback nəticəni *pisləşdirir*. Qazanc tək gözlərdən yox, yenidən təlimdən gəlir.
- Böyük və həlledici üstünlük **göstərmir**. Əksər ölçülərdə metod rəqiblərlə başabaşdır; “20× çox məlumatla öyrədilmiş modeli keçir” doğrudur, amma bir benchmarkda dolayı metriklərdə cüzi fərqlə.
- Ümumi bədii qabiliyyəti **nümayiş etdirmir**. Bu, kiçik sabit ayırdetmədə (224×224 piksel) ikon və sadə illüstrasiyalar çəkən səkkiz milyard parametrli tədqiqat modelidir, ümumi məqsədli dizayner deyil.
- Böyük ümumi modellə (GPT-5) müqayisə **eyni şərtlərdə deyil**: həmin model bu dar kodla-rəsm tapşırığı üçün qurulmayıb və optimallaşdırılmayıb; burada onu keçmək modellərin ümumi qabiliyyəti barədə az şey deyir.
- Bu, **xərclə gəlir**. Kətanı hər addımda render edib yenidən oxumaq bütün kodu bir kor keçiddə çıxarmaqdan yavaşdır — müəlliflər bunu qəbul edir.

Sübut nə qədər güclüdür

- **Öz şərtləri daxilində nəzarətli müqayisə üçün möhkəmdir.** Ablasiya testləri təmizdir: təlimi və ya yoxlamanı çıxarın, rəqəmlər düşür — deməli, müəlliflərin dediyi iki komponent həqiqətən işi görür.
- **Öz sürprizi barədə dürüstdür.** Sadəlövh vizual feedbackin *zərər verməsi* gizlədilmir və məqalənin ən maraqlı nəticəsidir — daha çox girişin həmişə daha yaxşı olması intuisiyasına faydalı düzəliş.
- **Leaderboard iddiasında nazikdir.** Daha çox məlumatla öyrədilmiş rəqiblər üzərində üstünlüklər kiçikdir və həmin rəqiblərdən birinin müəlliflərinin hazırladığı tək benchmarkdadır. Bir test dəstində dolayı metriklərdə kiçik fərqlər maraqlıdır, amma yekun söz deyil.
- **Miqyasda və real şəraitdə yoxlanmayıb.** Burada hər şey aşağı ayırdetmədə ikon və sadə illüstrasiyalardır. Eyni ideyanın mürəkkəb, yüksək ayırdetməli və real dizayn işlərində işləyib-işləməyəcəyi gələcək iş olaraq qalır — müəlliflər özləri də bunu deyir.

Niyə vacibdir

Bu məqalənin mərkəzindəki ideya demək olar utandıracaq qədər sadədir və rəsmdən xeyli uzağa gedir. Proqram kod yazaraq nəsə yaradacaqsa — veb səhifə, qrafik, diaqram, 3D səhnə — ya hamısını kor-koranə yazıb ümid edə bilər, ya da işlədikcə render edib istiqamətini düzəldə bilər. Bu dövrəni bağlamaq insan üçün ikinci təbiətdir; biz daim səhifəyə baxırıq. Bu modellər üçün isə təəccüblü dərəcədə yeni addımdır.

Məqaləni oxumağa dəyər edən yanında qoyduğu ulduz işarəsidir. Modelə görmək yolu vermək ona baxmağı öyrətməklə eyni deyil. Gözlər öyrədilməlidir; yalnız bundan sonra dövrə fayda verir — onda da qazanc real, amma ölçülüdür: tam başqa növ rəssam yox, daha sabit rəsmçi. Təsviri redaktə edilə bilən qrafikə çevirən alətlər — tətbiqlərin ikon və illüstrasiyalarının arxasına düşən növ — quranlar üçün sakit, praktik dərs faydalıdır. Buradakı qazanc daha çox məlumatdan yox, daha ucuz və ağıllı təlimdən gəldi. Bu, leaderboardda növbəti bir baldan daha yaxşı dərsdir.

Təmiz xülasə

Vektor qrafikası — əksər veb ikon və loqolarının arxasındakı redaktə edilə bilən, ölçüsü dəyişdirilə bilən kod — yaradan dil modelləri ənənəvi olaraq bunu “kor” edirdi: bütün rəsm əmrlərini nəticəni heç görmədən yazırdı. Guotao Liang-ın rəhbərlik etdiyi komanda Render-in-the-Loop təklif edir: hər addımdan sonra yarımçıq rəsmi render et və geri ver ki, model növbəti xətti kətana baxaraq çəksin. Onların əsas və dürüst nəticəsi budur ki, bunu hazır modelə sadəcə əlavə etmək onu *daha pis* edir; üstünlük yalnız model vizual feedbackdən

istifadə etməyi yenidən öyrəndikdən sonra yaranır və heç nə dəyişməyən xətləri atan rəsm-zamanı yoxlama buna kömək edir. Yenidən öyrədilmiş səkkiz milyard parametrli model standart benchmarkda iyirmi dəfəyə qədər çox məlumatla öyrədilmiş rəqiblərə çatır və ya az fərqlə keçir — həqiqi nəticədir, amma aşağı ayırdetmədə ikon və sadə illüstrasiyalar üçün dolayı ballarda kiçik fərqlərlə. Müəlliflərin vurğuladığı və yadda saxlamağa dəyər nəticə səmərəlilikdir: işinə yaxşı baxmaq sırf məlumat miqyasının bir hissəsini əvəz edə bilər.

No-BS yoxlaması

Məqalə nəyi göstərir: Vektor-qrafika modelini yarımçıq rəsmini addım-addım render edib ona baxmaq üçün yenidən öyrətmək, kor çəkməklə müqayisədə daha yaxşı və tamamlanmış nəticələr verir — daha az təlim məlumatı ilə bir standart benchmarkda daha çox məlumatlı rəqiblərlə rəqabətli və az fərqlə üstün nəticə.

Mümkündür, amma sübut olunmayıb: “İşinə baxmağın” ümumən məlumat miqyasını artırmaqdan daha yaxşı resept olması; eyni ideyanın mürəkkəb, yüksək ayırdetməli və real qrafikada kömək etməsi; benchmarkdakı kiçik fərqlərin insanın həqiqətən görə biləcəyi fərq olması.

Nəyi göstərmir: Vizual feedbackin öz-özünə kömək etdiyini (yenidən təlimsiz zərər verir); rəqiblər üzərində böyük və həlledici üstünlüyü; ümumi məqsədli rəsm qabiliyyətini; bu tapşırıq üçün qurulmayan GPT-5 kimi ümumi modellərlə tam ədalətli üz-üzə müqayisəni.

Əsas məhdudiyyətlər: Bir benchmark, qismən rəqib müəllifləri tərəfindən hazırlanıb; dolayı metriklərdə kiçik fərqlər; 8B model, 224×224 ölçüdə ikon və sadə illüstrasiyalar; hər addımı render edən dövrəyə görə daha yavaş generasiya.

Ümumi oxucu nə qədər əmin ola bilər? Dövrəni bağlamağın — çəkdikcə render edib yenidən baxmağın — həqiqətən kömək etdiyinə və bunun sadəcə açılmalı funksiya yox, öyrədilməli davranış olduğuna yüksək. Bu konkret modelin rəqibləri həlledici şəkildə keçdiyinə aşağı-orta. “20× çox məlumatı məğlub edir” cümləsini real, amma təvazökar, tək-benchmark nəticəsi kimi oxuyun. Yadda qalmalı ideyaya yüksək etibar: rəsm çəkən kod üçün işlədikcə baxmaq kor çəkməkdən yaxşıdır.

Mənbələr

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/