

06

KIJKEN

Een beeld wordt
een patchreeks

22

MAKEN

Waarom diffusie
begint met ruis

41

HERKOMST

Wat een
handtekening
bewijst

■ VAN SIGNAAL NAAR MODEL

AI die kijkt, hoort en maakt

Beeld, geluid en generatieve verbeelding ontleed.

Drie routes door beeld en geluid

Kies de route van herkennen, luisteren of maken. Dezelfde controlevraag keert terug: welke informatie zit werkelijk in het signaal?

ACHTERGROND

Zo lees je de voorbeelden

Een foto kan meetgegevens tonen; een generatieve afbeelding heet illustratie. Berekeningen volgen opgegeven getallen en eenheden. Een speelgoedmodel vereenvoudigt de werkelijkheid met zichtbare aannames. Fictieve scenario's onderzoeken een logische grens, terwijl onderzoeksresultaten hun eigen dataset en meettaak houden.

Bron: 26

+ KIJKOPDRACHT

Kies één foto uit dit nummer en maak twee lijstjes: wat direct zichtbaar is en wat je op basis van voorkennis vermoedt. Houd dat verschil bij de volgende artikelen in gedachten.



3-9

Kijken en meetpuzzels

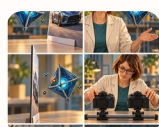
Strip en beeldmodellen op pagina 3-7; nieuwe stereoquiz op pagina 8; matchpuzzels op pagina 9.



10-15

Diepte en geluid

Eerste quiz en cijfers op pagina 10-11; dieptedossier op pagina 12-13; microfoon begint op pagina 14.



16-21

Frames en tegenbeelden

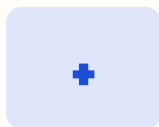
Audio op pagina 16; nieuwe strip en cijfers op pagina 17-18; bijschriften en golven op pagina 20-21.



22-27

Genereren en schalen

Diffusie op pagina 22-23; tijdquiz en schaalreferentie op pagina 24-25; GAN op pagina 27.



28-33

Vormen en verklaren

VAE op pagina 28; vensterpuzzels en tekenruimte op pagina 29-30; gedachten en video op pagina 31-32.



34-39

Afstanden en fouten

Nieuwe dekkingquiz en gedachten op pagina 34-35; rekenwerk op pagina 36; sensoren en beeldscores op pagina 38-39.



40-46

Pixels en praktijk

Modeldossier op pagina 40; herkomst, datasets en werkplaats op pagina 41-43; raadsels en papers op pagina 45-46.



47-50

Antwoorden en bronnen

Bestaande uitwerkingen op pagina 47; nieuwe puzzels en gedachten op pagina 48; nieuwe quizen op pagina 49; bronnen en vervolg op pagina 50.

VECTOR · VAN SIGNAAL NAAR MODEL

AXI en de gestreepte mok

Een model kan een handig patroon gebruiken dat buiten zijn oefenomgeving misleidt. De onderzoeker verandert de sticker, terwijl de mok dezelfde blijft.



1 Onderzoeker Wat zie je op deze camera-opname?



2 AXI Strepen! Mijn kaart wijst naar een kat.



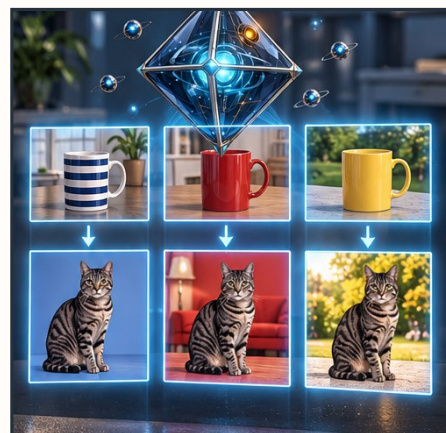
3 Onderzoeker Dan haal ik alleen de sticker van de mok.



4 AXI Mijn sterke aanwijzing is weg. De mok staat er nog.



5 Onderzoeker Test dezelfde vorm met andere kleuren en achtergronden.



6 AXI Een patroon dat goed scoort, moet ook op nieuwe gevallen passen.

Illustratie • AXI en de moktest zijn fictief; de shortcut is een didactisch voorbeeld.

Een aanwijzing kan een shortcut zijn

Als de trainingsvoorbeelden katten vaak met strepen tonen en mokken weinig variëren, kan textuur een opvallende aanwijzing worden. De strip

overdrijft dat mechanisme in een fictief laboratorium. Een nieuwe test varieert kleur, achtergrond en vorm afzonderlijk. Zo onderzoek je of de voorspelling het bedoelde object volgt of een toevallig meereizend patroon.



Een filter kijkt op veel plaatsen

Een convolutioneel netwerk gebruikt dezelfde kleine gewichten op verschillende plekken van een beeld. Dat deelt rekenregels in plaats van voor iedere pixel een nieuw patroon te leren.



FIG. 1 Foto • Water en rotsen leveren randen en textuur; de foto is geen meetgrafiek van een filter.

Van buurpixels naar een nieuw getal

Neem een fictief grijsbeeld met een rijwaarden $[0,0,10,10]$. Een eenvoudig verschilfilter vergelijkt aangrenzende waarden door rechts min links te nemen. Het levert $[0,10,0]$: de overgang valt op. Een echte convolutionele laag combineert meerdere pixels en kleurkanalen met geleerde gewichten. In veel software heet de bewerking convolution, al wordt het filter vaak zonder de wiskundige spiegeling als correlatie toegepast. Voor ons voorbeeld is de rekenafsprake direct gegeven.

Gedeelde gewichten zijn een aanname

Hetzelfde filter kan een verticale rand links én rechts vinden. Dat is nuttig wanneer een object op verschillende plaatsen verschijnt. Het netwerk kiest niet vooraf dat elke rand belangrijk is: training bepaalt welke filters voor het doel bruikbaar zijn. Niet-lineaire functies en meerdere lagen kunnen eenvoudige patronen combineren. Het effectieve kijkgebied groeit wanneer lagen verder verwerken of resolutie verminderen. Een lokaal patroon krijgt zo meer omgeving mee.

Een kaart is nog geen object

Een hoge filterrespons zegt dat de invoer op een geleerd patroon lijkt. Zij zegt niet automatisch “waterval”, “kat” of “deur”. Latere verwerking en het trainingsdoel bepalen welke labels volgen. De watervalfoto heeft heldere overgangen, textuur en bewegingsspooren. Je kunt daarin meerdere bruikbare patronen vinden zonder één naam als conclusie vast te leggen. Een classifier krijgt bovendien meestal een beperkte verzameling labels. Wat buiten die lijst valt, kan toch naar een bekende klasse worden geduwd. Vraag daarom welke categorieën mogelijk waren voordat je een score interpreteert.

SPEELGOEDFILTER

De regel rechts min links is hier expliciet gekozen. Hij demonstreert een lokale bewerking; hij beschrijft niet een daadwerkelijk geleerd filter van een bepaald model.



■ VECTOR · VAN SIGNAAL NAAR MODEL

Een residuele laag leert een aanvulling

Meer lagen bieden meer verwerking, maar maken optimalisatie moeilijker. Een residuele verbinding geeft een korte route naast een geleerde verandering.

Behoud een pad voor de invoer

Een residueel blok heeft schematisch uitkomst $y = x + F(x)$. De functie F leert een verandering, terwijl x via een korte verbinding mee gaat. Als een extra blok weinig nuttigs moet veranderen, kan een kleine F voldoende zijn. Bij verschillende dimensies kan een projectie nodig zijn om beide routes op te tellen. Het eenvoudige schema is geen belofte dat een volledig diep netwerk identiek gedrag kan behouden zonder andere effecten.

Waarom dit training helpt

De ResNet-studie onderzocht diepe netwerken met residuele bewerkingen en vond dat zij beter te optimaliseren waren dan vergelijkbare gewone stapelingen. De korte routes beïnvloeden hoe informatie en gradiënten door de berekening reizen. Daardoor kan grotere diepte bruikbaar worden. Het resultaat geldt voor de geteste architecturen en taken; “een extra laag helpt altijd” zou veel te sterk zijn. Dataset, optimizer, breedte en rekentijd blijven relevant.

Representatie en label zijn apart

Een beeldnetwerk kan eerst een representatie leren en daarna een classificatiekop gebruiken. Die kop vertaalt de representatie naar scores voor opgegeven labels. Je kunt dezelfde representatie soms met een andere kop op een andere taak inzetten. Dat heet transfer, maar vereist een toets op de nieuwe gegevens. Een netwerk dat op één verzameling objectfoto's goed scoort hoeft een microscopisch beeld niet te begrijpen. Verander je het domein, controleer dan de invoerconventies, labels en meetmethode. Een diepe representatie is geen universele visuele verklaring van alles wat de camera kan vastleggen.

ACHTERGROND

Een korte route is geen kortere taak

De verbinding $y = x + F(x)$ verandert de architectuur van het blok. Het aantal klassen, de kwaliteit van labels en de moeilijkheid van een nieuwe afbeelding veranderen daardoor niet automatisch. Goed trainen en goed toetsen blijven afzonderlijke stappen.

Bron: 3

DOSSIER MULTIMODALE AI

Van beeldvak tot gedeelde ruimte

Beeldverwerking en tekstverwerking kunnen vectoren leveren die gezamenlijk worden getraind. Overeenkomst blijft een modelsignaal dat je moet toetsen.



VECTOR · VAN SIGNAAL NAAR MODEL

Het beeld wordt een patchreeks

Een Vision Transformer knipt een beeld in vakken. Elk vak wordt naar een vector geprojecteerd en met andere vakken verwerkt.

Een telling die je kunt zien

Neem een beeld van 224×224 pixels en vierkante patches van 16×16 . Langs iedere rand passen 14 patches; samen zijn dat $14^2 = 196$ beeldpatches. Extra tokens voor classificatie tellen we in deze berekening niet mee. Verdubbel beide beeldranden naar 448 pixels bij dezelfde patchmaat: $28^2 = 784$ patches. De beeldoppervlakte en het aantal patches zijn verviervoudigd.

Een patch is een vector, geen woord

De pixels en kleurkanalen van een patch worden als invoer naar een geleerde projectie gebruikt. Positie-informatie geeft aan waar het vak ligt. Attention verbindt vervolgens representaties van verschillende vakken. Het systeem kan zo informatie over verder uit elkaar liggende delen

mengen. De term token zegt alleen dat het element in een reeks wordt verwerkt. Een patchtoken heeft niet vanzelf de betekenis “oog” of “wiel”, en zijn grens hoeft niet met een objectgrens samen te vallen.

De afweging verschuift

Kleinere patches bieden meer lokale details maar produceren meer elementen. Bij dense attention groeit het aantal paren tussen elementen kwadratisch. Als 196 patches 784 worden, wordt 196^2 naar 784^2 zestien keer zo groot. Dat is een telling van mogelijke paren, geen exact zestienvoudige looptijd van het complete systeem. Andere onderdelen, geheugenverkeer en speciale attentionvormen beïnvloeden de rekening. Het ViT-onderzoek liet zien dat een transformer op grote beeldcollecties sterke classificatierepresentaties kan leren. Dat vervangt geen afzonderlijke test op kleine objecten, nieuwe achtergronden of onbekende domeinen.

REKENAFSPRAAK

De patchtelling veronderstelt dat 16 exact in de beeldrand past. We tellen alleen beeldpatches, geen eventuele extra tokens of implementatiebuffers.



VECTOR · VAN SIGNAAL NAAR MODEL

CLIP leert beeld en tekst vergelijken

Twee encoders maken vectoren: één uit een afbeelding, één uit tekst. Contrastief leren trekt passende paren bij elkaar ten opzichte van andere paren.

De batch levert de vergelijking

Stel dat een minibatch foto's van een mok, een fiets en een boom bevat, elk met een bijbehorende beschrijving. Het trainingsdoel verhoogt de overeenkomst van de passende paren ten opzichte van de niet-passende paren in de batch. De tekst geeft natuurlijke taal als leerhulp, zonder dat ieder beeld één vaste categorie hoeft te krijgen. Het oorspronkelijke CLIP-onderzoek gebruikte 400 miljoen verzamelde beeld-tekstparen. Dat is een specifieke onderzoekscollectie, geen volledige beschrijving van alle latere modellen.

Een label is een kandidaattekst

Voor classificatie kun je beschrijvingen van mogelijke categorieën omzetten in tekstvectoren.

Het beeld krijgt een beeldvector; overeenkomst bepaalt welke kandidaat hoger scoort. Verander je “een foto van een mok” in een minder passende of afwijkende formulering, dan kan de vergelijking veranderen. De aangeboden categorieën blijven een keuze van de meetopstelling. Als geen kandidaat het beeld goed beschrijft, kan er toch één bovenaan eindigen. Een hoge relatieve score hoeft geen volledige zekerheid over de wereld te zijn.

Gedeelde ruimte, verschillende vragen

CLIP is in de basis een vergelijkingsmodel, geen zelfstandig systeem dat alle objecten opsomt of een nieuw beeld tekent. Zijn representaties kunnen onderdelen van grotere systemen worden. Een beeldbijschrift, een generator en een retriever gebruiken zulke onderdelen met andere doelen. Houd die rollen uit elkaar. De cosinus tussen [3,4] en [4,3] is 0,96, maar zelfs een perfecte overeenkomst tussen twee geleerde vectoren bewijst geen exacte feitelijke gelijkheid van hun bronnen. De interpretatie hangt af van training, kandidaten en de concrete taak. ♦



AXI De beste kandidaat wint een vergelijking. Welke kandidaat ontbrak er misschien in de lijst?

Quiz: van twee pixels naar één afstand

Vier vragen blijven binnen de ideale stereo-afpraak. Alle getallen zijn zelfgekozen; antwoorden met berekening staan achterin.

Bewaar dezelfde camera

Gebruik $f = 800$ pixels en $b = 0,10$ m, tenzij de vraag die expliciet verandert. De camera-afpraak staat op pagina 12–13; uitwerkingen op pagina 49.

1 Bij $d = 20$ pixels hoort welke diepte?

- ☐ (A) 4 m
- ☐ (B) 0,25 m
- ☐ (C) 160 m

Reken hier

2 Bij dezelfde camera wordt $d = 10$ pixels. Wat is Z ?

- ☐ (A) 2 m
- ☐ (B) 4 m
- ☐ (C) 8 m

Reken hier

3 Hetzelfde punt blijft op 4 m; b wordt 0,20 m. Wat wordt d ?

- ☐ (A) 10 pixels
- ☐ (B) 40 pixels
- ☐ (C) 20 pixels

Reken hier

4 Een egaal beeldgebied heeft meerdere even goede matches. Wat is een verantwoord resultaat?

- ☐ (A) De mooiste afstand als exact presenteren.
- ☐ (B) Alleen de kleur als meetbewijs gebruiken.
- ☐ (C) De match als onzeker of ongeldig markeren.

Reken hier

Jouw antwoorden

1 2 3 4

De match die wél uniek is

Drie kleine beeldpuzzels lossen op zonder foto of software. Alle regels staan in de opgave; bewijzen staan op pagina 48.

1 Op één gerectificeerde beeldrij heeft een punt links coördinaat 42. Het bijbehorende patroon verschijnt rechts uitsluitend op 22; andere patronen zijn verschillend. Wat zijn d en Z bij $fb=80$ pixel·m?

Mijn antwoord

2 Een linker patroon ABA kan rechts matchen met ABA op coördinaat 12 of ABA op 22. Beide zijn volgens de opgegeven regel toegestaan. Is de disparity uniek bekend?

Mijn antwoord

3 Een perfect stereoapparaat heeft $fb=60$ pixel·m. Drie punten hebben dieptes 2,3,6 m. Welke unieke disparities horen erbij?

Mijn antwoord

Het patroon is de meetinformatie

Een match wordt alleen toegestaan als het aangegeven patroon gelijk is. Afstanden volgen daarna uit de al beschreven camerageometrie.

Quiz: wat leert het beeldmodel?

Zes vragen scheiden een pixelbewerking, representatie, taak en meetresultaat. Kies de uitspraak die precies door de gegeven aannames wordt gedragen.

Verschillende onderdelen

Een filterrespons is nog geen objectnaam. Een patchtelling is nog geen looptijd. Een labelscore is nog geen volledige beschrijving. Houd bij ieder antwoord de grootte en de taak zichtbaar. De toelichting staat op pagina 47; de volledige voorbeelden vind je op pagina 4–7, 19–20, 22–23, 27–28, 32–33, 37.

- 1 Het verschilfilter rechts min links werkt op de rij [0,0,10,10]. Wat levert het tussen aangrenzende waarden?**
- (A) [0,10,0], een sterke overgang in het midden.
 - (B) [0,0,10,10], altijd dezelfde invoer.
 - (C) Het objectlabel “waterval”.

- 2 Een residueel blok schrijft $y = x + F(x)$. Welke betekenis past daarbij?**
- (A) x wordt uit de berekening verwijderd.
 - (B) Iedere extra laag garandeert betere classificatie.
 - (C) Een korte invoerroute wordt gecombineerd met een geleerde verandering.

- 3 Een 224×224 -beeld gebruikt 16×16 -patches. We tellen geen extra tokens. Hoeveel beeldpatches zijn er?**
- (A) 14.
 - (B) 196.
 - (C) 224.

4 Wat doet CLIP in de beschreven basisopzet?

- (A) Het vergelijkt geleerde beeld- en tekstrepresentaties.
- (B) Het meet direct of iedere afgebeelde gebeurtenis waar is.
- (C) Het tekent zelfstandig een volledig nieuw beeld uit iedere prompt.

5 Een bijschrift noemt een glas dat niet zichtbaar is. Welke fout past bij deze beschrijving?

- (A) Alleen een verkeerde patchtelling.
- (B) Alleen een onjuist aantal cameraframes.
- (C) Een objecthallucinatie in de gegenereerde beschrijving.

6 Geen trainingscamera keek achter een kast. Wat mag je over een overtuigend nieuw gerenderd beeld van die verborgen kant concluderen?

- (A) Het is per definitie een exacte meting van de verborgen kant.
- (B) Het kan een plausibele invulling zijn die niet uniek door de foto's is bepaald.
- (C) Iedere radiantieveldrepresentatie kent automatisch alle onzichtbare oppervlakken.

Jouw antwoorden

1	2	3
4	5	6

■ VECTOR · VAN SIGNAAL NAAR MODEL

Zes getallen achter beeld en geluid

Sommige cijfers beschrijven ruwe data, andere een gepubliceerde trainingscollectie. De betekenis en conventie staan er steeds bij.

DATAPUNT

1,92 MB

één minuut van de beschreven ruwe mono-audio

Een minuut geluid zonder compressie

Mono-audio met 16.000 samples per seconde en zestien bits per sample gebruikt 32.000 bytes per seconde. Eén minuut is 1.920.000 bytes, oftewel 1,92 MB in decimale notatie, zonder header. Meer talen of moeilijker woorden veranderen die vaste bestandstelling niet.

DATAPUNT

784

beeldpatches bij 448×448 en patchrand 16

De patchreeks groeit met oppervlakte

Bij 16×16 -patches heeft 224×224 pixels 196 vakken; 448×448 heeft 784. Dubbele randen geven viermaal het aantal patches. Densere paren tussen patches gaan van 38.416 naar 614.656: zestienmaal zoveel mogelijke vergelijkingen.

DATAPUNT

1,493 GB/s

de berekende onbewerkte videostroom

Ruwe video heeft een flinke stroom

Voor 3840×2160 pixels, drie achtbitkanalen en 60 frames per seconde is de ruwe stroom $3840 \times 2160 \times 3 \times 60 = 1.492.992.000$ bytes per seconde. Dat is circa 1,493 GB/s zonder compressie, audio, header of andere overhead.

DATAPUNT

400 miljoen

beeld-tekstparen in het oorspronkelijke CLIP-onderzoek

CLIP gebruikte heel veel paren

Het oorspronkelijke CLIP-onderzoek beschreef 400 miljoen beeld-tekstparen voor de trainingsopzet. Eén paar geeft een tekstuele aanwijzing bij een beeld, maar geen garantie dat die tekst ieder object nauwkeurig beschrijft. Datakwaliteit en selectie blijven naast omvang relevant.

DATAPUNT

64×

minder ruimtelijke posities in het latente voorbeeld

Een latente ruimte verandert de telling

Als een gekozen beeldencoder de rand van 512 naar 64 elementen terugbrengt, daalt het aantal ruimtelijke posities van 262.144 naar 4096: een factor 64. Het aantal kanalen en bytes per getal kunnen anders zijn. Dit is dus geen automatische 64-voudige bestandcompressieclaim.

DATAPUNT

680.000 uur

supervisie in de oorspronkelijke Whisper-studie

Whisper beschreef 680.000 uur

De oorspronkelijke Whisper-studie gebruikte 680.000 uur aan meertalige, multitask-supervisie. Die omvang maakte de onderzochte brede opzet mogelijk. Zij betekent niet dat elke taal evenveel voorbeelden kreeg of dat iedere onbekende opname dezelfde foutkans heeft. Kijk daarvoor naar afzonderlijke evaluaties.



Twée camera's geven een driehoek

Een beeld bevat richtingen. Twee gekalibreerde camera's kunnen daarbij een afstand helpen bepalen. Volg een expliciet gekozen meetgeometrie.



FIG. 1 AI-illustratie • een stereo-opstelling met kaart en miniatuur; de camera's staan naast elkaar, zonder exacte kalibratiematen. AXI en de onderzoeker zijn fictief.

De afspraak van parallelle camera's

Neem twee ideale pinholecamera's met parallelle assen, onderlinge afstand b en dezelfde brandpuntsafstand f in pixels. Hun beelden zijn gerectificeerd: overeenkomstige punten liggen op dezelfde beeldrij. Voor een punt op diepte Z verschillen de horizontale coördinaten $d = fb/Z$ pixels. Dit is de disparity. De formule volgt uit gelijkvormige driehoeken: de camera-afstand en de beeldverschuiving meten dezelfde verhouding.

Van pixels naar meter

Ons fictieve camerapaartje heeft $f = 800$ pixels en $b = 0,10$ m. Een betrouwbaar gematcht punt heeft $d = 20$ pixels. Dan $Z = 800 \times 0,10 / 20 = 4,00$ m. De pixel-eenheden vallen weg. Twee andere punten met $d = 40$ en $d = 10$ liggen onder precies dezelfde aannamen op 2,00 en 8,00 m. Een kleinere disparity hoort hier bij grotere diepte.

De kwetsbare stap zit vóór het delen

Een algoritme moet eerst in beide beelden hetzelfde wereldpunt vinden. Een effen muur levert weinig unieke patronen; een voorgrondobject kan een achtergrondpunt in één beeld bedekken. AI kan zulke overeenkomsten leren zoeken, maar een plausibele match is niet automatisch de juiste. Bewaar een geldigheidsmasker: “geen betrouwbare match” is een informatieve uitkomst. De volgende pagina onderzoekt wat één pixel verschil met onze vier meter doet.

IDEALE MINIATUUR

Alle camera-afmetingen en afstanden in de rekenvoorbeelden zijn aangenomen. Werkelijke camerakalibratie kan extra onzekerheid bevatten.



■ VECTOR • ONDERZOEK MET AANNAMEN

Eén pixel maakt de afstand scheef

De inverse relatie tussen verschuiving en diepte maakt een symmetrische pixelfout tot een asymmetrisch afstandsinterval.

1 Dezelfde rekenlijn, andere match

Houd $fb = 80$ pixel·m vast. Bij $d = 19,20,21$ krijgen we $Z \approx 4,2105; 4,0000; 3,8095$ m. De verschillen rond vier meter zijn $+0,2105$ en $-0,1905$ m. Een band van één pixel is dus niet automatisch een symmetrische band in meter. Bij kleine disparities kan hetzelfde pixelverschil nog meer afstand schelen.

2 Geleerde diepte uit één beeld

Dense Prediction Transformers combineert transformerrepresentaties met een decoder voor onder meer monoculaire diepteschatting. Zo'n geleerd model gebruikt patronen uit trainingsbeelden. Voor onze fictieve relatieve variant leggen we vast dat zij alleen diepteverhoudingen geeft; pas een aparte schaalreferentie maakt meters mogelijk. Dat is een contract van dit voorbeeld, geen uitspraak dat ieder bestaand monoclair model uitsluitend relatieve diepte kan leveren.

3 Een benchmark heeft een meetomgeving

De oorspronkelijke KITTI-stereobenchmark uit 2012 gebruikte gekalibreerde camera's en een laserreferentie in echte verkeersomgevingen. Het onderzoek benoemt onder meer reflecties, transparantie en afdekking als lastige gevallen. Een goede score op zo'n taak beoordeelt die verzameling en foutregel. Onze platte fotokaart uit de strip vraagt een eigen proef: vergelijk gematchte disparity met een meetreferentie, en rapporteer ook welke pixels als ongeldig zijn gemarkeerd.

DATAPUNT

19 px

aangenomen onderste disparity

DATAPUNT

4,2105 m

bij 19 pixels

DATAPUNT

3,8095 m

bij 21 pixels

DATAPUNT

2012

oorspronkelijke KITTI-stereotaak



VECTOR · VAN SIGNAAL NAAR MODEL

Een microfoon levert tijdwaarden

Voordat een spraakmodel woorden voorspelt, wordt een veranderend geluid in een digitale reeks vastgelegd. Sampling en filtering bepalen wat die reeks bewaart.

Van geluid naar modelinvoer

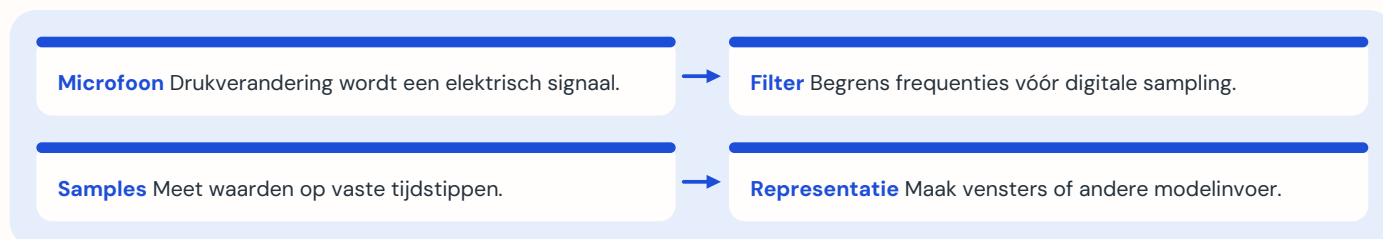


FIG. 1 Een schematische meetketen; details verschillen per systeem.

Van druk naar samples

Een microfoon zet drukverandering om in een elektrisch signaal. Een digitaal systeem meet dat signaal op opeenvolgende tijdstippen. Bij 16.000 samples per seconde is de tussenafstand 1/16.000 seconde, dus 62,5 microseconden. Een zestienbitwaarde geeft 65.536 mogelijke codes, afhankelijk van de gekozen schaal. Dat is een discretisatie van het signaal, geen rechtstreeks alfabet van klanken. Hetzelfde gesproken woord kan met verschillende stemmen en snelheden heel andere samples opleveren.

Hoge frequenties kunnen verkeerd terugkomen

Onder de ideale aanname van een bandbegrensd signaal vraagt reconstructie een samplingfrequentie groter dan tweemaal de hoogste frequentie. In de praktijk gebruikt de meetketen een antialiasfilter vóór sampling. Een cosinus van 10 kHz bemonsterd op 16 kHz kan dezelfde samplewaarden geven als een cosinus van 6 kHz: de hogere frequentie vouwt terug. Meer slimme interpretatie achteraf vertelt niet uniek welke van die twee oorspronkelijk aanwezig was. Daarom is de fysieke invoerketen onderdeel van de bewijsvoering.

Een spectrogram is een nieuwe weergave

Een spectrogram verdeelt de tijdreeks in vensters en geeft per venster frequentie-informatie weer. De vensterlengte bepaalt een afweging tussen

tijd- en frequentiedetail. Mel-schaalverwerking groepeerd frequenties op een andere schaal dan een gewone lineaire as. Het spraakmodel ziet vervolgens representaties die aan deze voorbewerking zijn verbonden. Vraag bij een vergelijking dus welke sampling, filtering, vensters en normalisatie zijn gebruikt. Een mooi transcript begint met een specifieke meting, niet met een onbeperkt oor.

DATAPUNT

62,5 μ s

tijd tussen samples bij 16 kHz



VECTOR · VAN SIGNAAL NAAR MODEL

Een transcript combineert horen en voorspellen

Een spraakmodel gebruikt een gemeten signaal en geleerde taalpatronen. Dat helpt bij onduidelijk geluid, maar kan ook een plausibel ontbrekend stuk invullen.

Wat doet een encoder-decoder?

Whisper gebruikt een encoder die audiorepresentaties verwerkt en een decoder die tokens produceert. De decoder kan eerdere teksttokens benutten, zodat het transcript taalkundig samenhangt. De training combineerde verschillende talen en spraaktaken. Het oorspronkelijke onderzoek beschreef 680.000 uur aan verzamelde supervisie. De dataomvang vertelt niet dat iedere taal, stem of opnameomstandigheid even goed is vertegenwoordigd. Daarvoor moet je resultaten per scenario bekijken.

Waarom helpt context?

Een zachte klank in “ik betaal met de ...” kan door de rest van de zin beter te herkennen lijken. Taalpatronen geven een voorafverwachting over mogelijke woorden. Als het geluid weinig informatie bevat, kan die verwachting te zwaar wegen. Een transcript kan dan een echt klinkende maar niet gehoorde naam of zin bevatten. Controleer zulke details aan de opname. Een automatisch toegevoegd leesteken of hoofdletter is bovendien een modelkeuze, geen geluid dat de microfoon letterlijk heeft gemeten.

Hoe toets je het resultaat?

Leg eerst vast wat als fout telt: vervanging, ontbrekend woord, extra woord, naam of tijdstempel. Word error rate gebruikt vervangingen, weglatingen en invoegingen ten opzichte van een referentie, maar één gemiste ontkenning kan inhoudelijk zwaarder wegen dan een klein functiewoord. Bewaar daarom ook voorbeelden en betekenisvolle fouten. Vergelijk heldere spraak, achtergrondgeluid en stilte met dezelfde instellingen. Een model dat op één benchmark goed scoort kan in een stille passage alsnog woorden produceren. De tekst is een voorspelling op basis van geluid; bij kritische details blijft de opname de controlebron.

CONTROLEPUNT

Luister bij een naam, getal of ontkenning naar het originele fragment. Noteer of de informatie hoorbaar is, alleen uit context wordt afgeleid of helemaal ontbreekt.

VECTOR · VAN SIGNAAL NAAR MODEL

CTC koppelt lange tijd aan korte labels

Een audiosignaal heeft veel tijdstappen; een woord heeft weinig letters. Connectionist Temporal Classification kan mogelijke alignments samen nemen zonder iedere lettertijd apart te labelen.

Wat doet de lege code?

CTC voorspelt op tijdstappen labels plus een blank, een aparte lege code. Opeenvolgende herhalingen van hetzelfde label worden samengevoegd; daarna worden blanks verwijderd. De volgorde is essentieel. De padreeks L, L wordt één L. De reeks L, blank, L wordt LL: de blank scheidt twee voorkomens voordat hij verdwijnt. Het blank is dus geen spatie in het uiteindelijke woord en geen gewone letter die betekenis draagt.

Meerdere paden voor één uitkomst

Een korte labelreeks kan uit veel tijdpaden ontstaan. CTC telt tijdens training de kansen van passende paden samen. Daardoor hoeft de dataset niet bij iedere sample de exacte lettergrens aan te wijzen. Het basismodel neemt een factorisatie van padkansen over tijdstappen aan; de gebruikte netwerkrepresentaties kunnen toch context uit het signaal bevatten. Een apart taalmodel of andere decoder kan bij de uiteindelijke keuze aanvullende structuur gebruiken. De alignmentmethode en het complete herkenningssysteem zijn niet hetzelfde.

Leren zonder elk transcript?

Wav2vec 2.0 leert eerst spraakrepresentaties met een zelfgesuperviseerd doel op audio, en past daarna aan met gelabelde gegevens. Zo worden ongelabelde signaalpatronen bruikbaar voordat ieder fragment een transcript heeft. Het maakt labels niet overbodig voor iedere eindtaak, en het is een andere trainingsopzet dan Whisper. Vergelijk daarom niet alleen datagrootte: vraag welke doelen, labels en evaluatiesets zijn gebruikt. Een goede representatie, een correcte alignment en een bruikbaar transcript vormen

Leren zonder elk transcript?

drie verbonden, maar afzonderlijk toetsbare onderdelen.

REKEN NA

Pas eerst herhalingsreductie toe en verwijder daarna blanks: L blank L wordt LL; L L blank wordt L. Dezelfde drie tijdstappen kunnen dus een andere labelreeks geven.

BEGRIPPEN Tijd en label

De labelreeks is korter dan het tijdpad. Houd de codes en de uiteindelijke tekst uit elkaar.

Tijdpad De labels en blanks op opeenvolgende signaaltijdstappen.

Labelreeks De kortere uitkomst na herhalingsreductie en blankverwijdering.

Bron: 9

VECTOR · ONDERZOEK MET AANNAMEN

AXI en de auto zonder achterkant

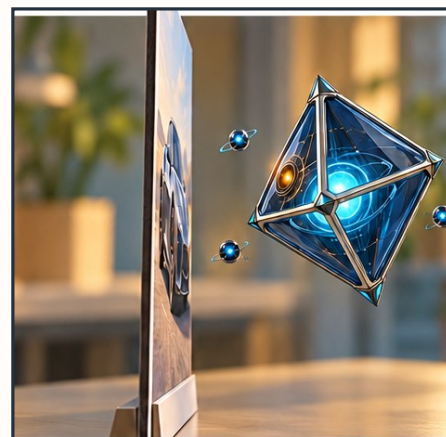
Een geloofwaardig beeld laat AXI een ruimte veronderstellen die de kaart niet bevat.



1 **AXI** Een enorme auto. Dat wordt een diepe garage!



2 **Onderzoeker** Loop eens één stap naar rechts.



3 **AXI** De auto is een platte kaart.



4 **Onderzoeker** Dit zijn twee camera's. Ze meten het oppervlak.



5 **AXI** De miniatuur heeft andere punten op andere dieptes.



6 **Onderzoeker** Een geloofwaardige afbeelding en een ruimtelijk object krijgen verschillende toetsen.

AI-illustratie • Zes fictieve scènes met AXI en een onderzoeker. Dialoog staat onder ieder kader; de beelden leveren geen meetgegevens.

Een andere kijkhoek voegt informatie toe

De strip is een fictieve demonstratie. De twee camerastanden bekijken de fysieke kaart; haar

afgebeelde wereld is iets anders dan de diepte van het oppervlak waarop de inkt zit. Een dieptemeting moet zeggen welke van die twee vragen wordt gesteld.

Zes cijfers tussen geluid en diepte

Nieuwe telweetjes gebruiken aangenomen audio-instellingen en expliciete meetregels. Onderzoeksnamen zeggen waar de methode vandaan komt.

01 32.000 · 1 · Twee seconden in samples

samples in dit fragment. Bij aangenomen 16 kHz heeft twee seconden audio 32.000 samples. Dat is nog geen aantal gesproken woorden: dezelfde tijdreeks kan verschillende stiltes, snelheden en sprekers bevatten.

02 25 ms · 2 · Een venster van 25 ms

gekozen vensterduur. 400 samples / 16.000 samples per seconde = 0,025 s = 25 ms. Een frame kan meerdere samples bevatten; noem beide aantallen zodat een tijdsberekening reproduceerbaar wordt.

03 4 · 3 · Vier volledige vensters

volledige vensters. Bij 1000 samples, venster 400 en stap 200 zijn starts 0,200,400,600 geldig. Er zijn vier volledige vensters zonder padding. Start 800 zou buiten het fragment eindigen.

04 2020 · 4 · Ongetranscribeerd is niet ongeleerd

wav 2vec 2.0–paper. wav 2vec 2.0 maskeert latente spraakrepresentaties en gebruikt een contrastieve leertaak. De methode leert eerst uit audio en kan daarna op transcripties worden aangepast. “Zelfsupervisie” benoemt de geconstrueerde leertaak, geen afwezigheid van trainingskeuzes.

05 ×2 · 5 · Bij halve disparity dubbele diepte

diepte bij halve disparity. Onze ideale stereo heeft $fb=80$ pixel·m. $d=20$ geeft 4 m, $d=10$ geeft 8 m. De inverse relatie heeft een meetcontract nodig: betrouwbare match, bekende geometrie en een geschikte scène.

06 194 / 195 · 6 · 194 en 195

KITTI stereo 2012. De oorspronkelijke KITTI–stereotaak uit 2012 had 194 trainings– en 195 testbeeldparen. Dat jaartal hoort erbij; andere KITTI–taken en latere versies hebben andere aantallen en regels.

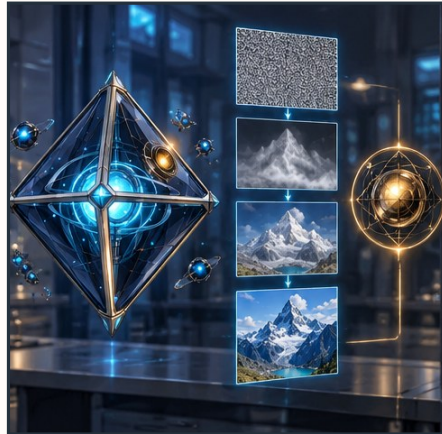
VECTOR · VAN SIGNAAL NAAR MODEL

AXI en het geldige zegel

Een synthetisch bergbeeld kan een geldige herkomstverklaring hebben. Dat bewijst iets over het bestand en zijn maker, maar vraagt nog een andere vraag over de scène.



1 Onderzoeker Deze bergen lijken gefotografeerd. Wat bewijst het zegel?



2 AXI De ondertekende herkomstclaim hoort bij dit bestand.



3 Onderzoeker Deze afbeelding is geen opname vanuit dit raam. Hoe is het bestand ontstaan?



4 AXI Het geregistreerde maakproces noemt een generator.



5 AXI De handtekening kan geldig zijn, juist omdat die oorsprong eerlijk is vastgelegd.



6 Onderzoeker Dan bewaren we herkomst en bewijs van een gebeurtenis als twee vragen.

Illustratie · AXI onderzoekt een fictief synthetisch beeld; de strip verbeeldt het verschil tussen herkomst en gebeurtenis.

Twee controles naast elkaar

Een digitale handtekening kan de binding van herkomstclaims aan een bestand en de ondertekenaar controleerbaar maken. De claim kan eerlijk zeggen dat het beeld is gegenereerd. Een

geldig zegel verandert de afbeelding dan niet in een camerawaarneming van die bergen. De strip speelt in een fictief laboratorium: de onderzoeker bewaart herkomst en waarneming in afzonderlijke dossiers.

VECTOR · VAN SIGNAAL NAAR MODEL

Een bijschrift kan te veel zien

Een beeldbijschrift is gegenereerde taal die op beeldinformatie is geconditioneerd. De taal kan objecten noemen die wel passen bij het verhaal, maar niet in de pixels staan.

AXI: Een glas past bij een tafelverhaal. Maar staat het ook op deze foto?

Onderzoeker: Controleer de pixels voordat je het verhaal aanvult.

Een gebruikelijk tafereel trekt details aan

Een foto van een eettafel kan een model aanzetten tot “borden, glazen en bestek”, ook wanneer een glas ontbreekt. Het systeem heeft geleerd welke woorden vaak samen met zulke scènes verschijnen. Het beeld levert aanwijzingen; de decoder heeft ook een taalverdeling. Object Hallucination in Image Captioning onderzocht zulke extra objecten en stelde een maatstaf voor naast gewone captionkwaliteit. Een bijschrift dat goed klinkt kan dus op één aparte, feitelijke eigenschap falen.

Noem wat zichtbaar is

Een controle begint met concrete claims: staat er een glas, hoeveel personen zie je, is het bord leesbaar? Andere uitspraken, zoals “de mensen vieren een verjaardag”, kunnen aannames over aanleiding of intentie zijn. Een beeld kan taart tonen zonder de gelegenheid uniek vast te leggen. Segmentatie helpt een gebied afbakenen, maar bewijst niet vanzelf welk object of welke gebeurtenis het gebied voorstelt. Een maskerrand is een ruimtelijk resultaat, geen compleet semantisch verslag.

Verander een detail zorgvuldig

Voor een eigen test kun je een voorwerp uit een foto afdekken en vragen of het bijschrift het nog noemt. Behoud dezelfde vraag en vergelijk beide uitkomsten. Afdekken verandert ook de zichtbare context, dus rapporteer precies wat je hebt gedaan. Een crop kan nuttig zijn om een regio te onderzoeken, maar sluit de rest van de scène uit. Het doel is niet één dramatische fout vinden: het is meten welke informatie de tekst daadwerkelijk uit het aangeboden beeld volgt. Die controle geldt

ook voor mooie beschrijvingen van wetenschappelijke afbeeldingen. 

SEGMENTATIE

Afbakenen is een andere taak

Een omtrek en een betekenis beantwoorden verschillende vragen.

Een bruikbaar masker

Een segmentatiemodel kan een gebied bij een punt of vak afbakenen. Test de rand tegen de gewenste pixels. Als het systeem een achtergrondgebied selecteert, hoeft dat geen taalhallucinatie te zijn: het is een fout in een andere taak met een andere maatstaf.

Bron: 27

■ VECTOR • VAN SIGNAAL NAAR MODEL

Twée signalen kunnen elkaar opheffen

De simulatie mengt twee ideale sinusgolven met gelijke amplitude en frequentieparameter. Fase bepaalt of de som hoger of lager wordt.

Drie fases bij $A = 1,5$

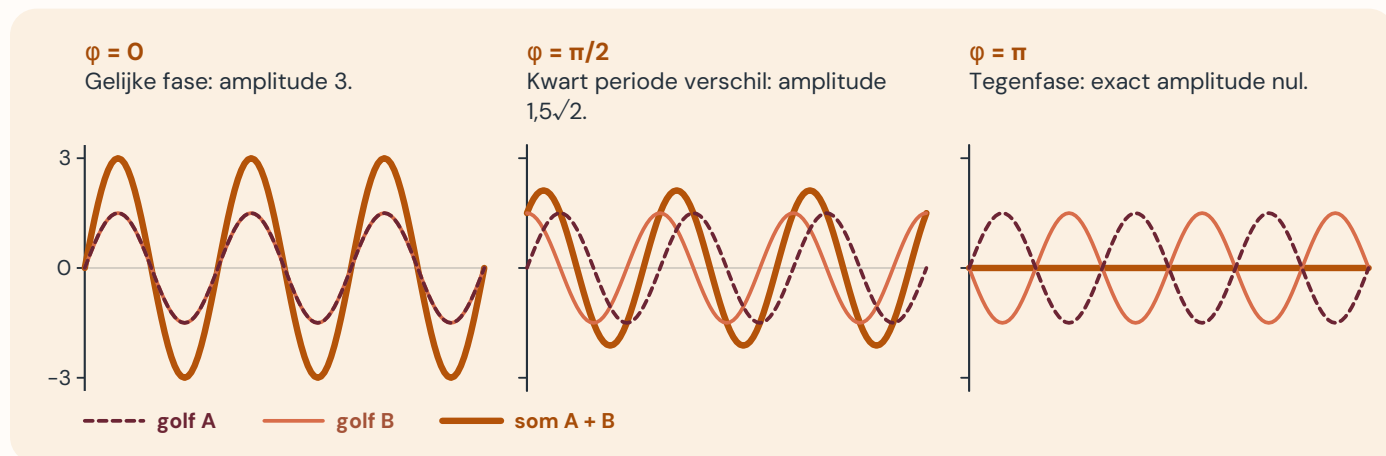


FIG. 1 Ideale, berekende amplitudes: geen gemeten geluidsterkte.

De formule geeft een voorspelling

We gebruiken $y_1 = A \sin(fx)$ en $y_2 = A \sin(fx + \varphi)$. x en f zijn dimensieloos; φ staat in radialen. De som wordt $2A \cos(\varphi/2) \sin(fx + \varphi/2)$. De amplitude is dus $2A \cdot |\cos(\varphi/2)|$. Met $A = 1,5$ en $\varphi = 0$ is zij 3. Bij $\varphi = \pi/2$ is zij $1,5\sqrt{2} \approx 2,121$. Bij $\varphi = \pi$ is zij nul. De schuifregelaar rondt φ af, dus 3,14 geeft een heel kleine rest in plaats van exacte nul.

Fase en superpositie

$y_1 = A \sin(fx)$, $y_2 = A \sin(fx + \varphi)$. x en f zijn dimensieloos; φ is in radialen.

Wat frequentie hier wel verandert

Een grotere f geeft meer perioden langs dezelfde x -as, maar bij hetzelfde faseverschil blijft de berekende somamplitude gelijk. Het model heeft geen verschillende frequenties, damping of microfoonruis. Het laat dus geen zwevingen zien en geeft geen gemeten geluidsniveau in decibels. De demonstratie is geschikt om fase en superpositie apart te onderzoeken. Wil je werkelijke audio reconstrueren, dan komen sampling, kalibratie en de fysieke meetpositie erbij.

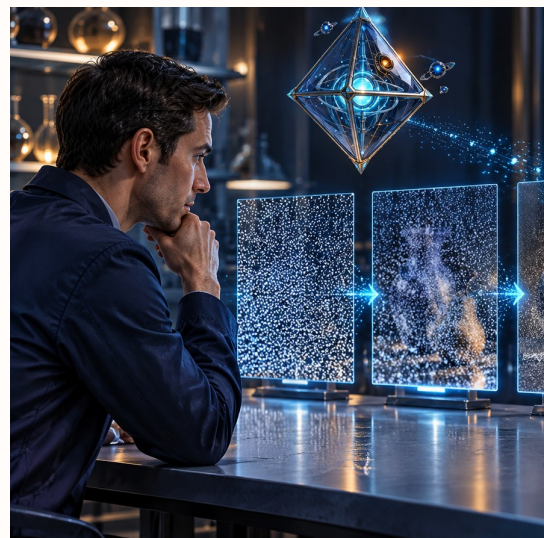
Waarom mengen informatie kan verliezen

Als twee tegengestelde signalen bij één punt exact optellen tot nul, kun je uit die som alleen niet bepalen welke twee bijdragen er waren. Ook twee nulbijdragen zouden dezelfde som geven. Dat is een inverse ambiguïteit, geen magie. Bij fysieke golven moet je bronnen, andere plaatsen en energiestromen samen beschrijven. Deze algebra volgt alleen de uitwijking op de gekozen as. Voorspel eerst de drie amplitudes en vergelijk ze met de curves. Bewaar het verschil tussen een ideaal model en een echt geluidsveld.

DOSSIER MULTIMODALE AI

Van ruis naar een nieuw beeld

De afbeelding verbeeldt vijf fasen van een generatief proces. De tekst onderscheidt de gecontroleerde trainingsruis van de gekozen samplingroute.



VECTOR • VAN SIGNAAL NAAR MODEL

Diffusie leert een ruisstap omkeren

Een diffusie-opzet beschadigt trainingsbeelden geleidelijk met ruis. Het model leert informatie waarmee een omgekeerd proces nieuwe samples kan maken.

De heenweg is gecontroleerd

In een eenvoudige formulering wordt een bekend beeld gemengd met Gaussische ruis volgens een ruisplanning. Het tijdstapnummer vertelt hoe sterk de beschadiging is. Tijdens training weet het systeem welk oorspronkelijk beeld en welke ruis zijn gebruikt. Het kan een netwerk bijvoorbeeld laten voorspellen welke ruis aan het huidige mengsel bijdroeg. Er bestaan ook andere parametrisaties. Het gegeven trainingsdoel is dus preciezer dan “het model leert een wazig beeld scherp te maken”.

De terugweg heeft een geleerd patroon

Bij generatie begin je met een ruissample en neem je stappen volgens een gekozen sampler en geleerd model. De geschatte richting of ruisinformatie stuurt een volgende toestand. Het

nieuwe beeld hoeft niet één achterliggend origineel te hebben dat ergens verstopt zat. De procedure genereert een sample uit geleerde structuur. Het aantal stappen, de planning en eventuele toevalscomponenten horen bij de uitvoering. Een tussenbeeld is geen moment waarop het systeem een voorwerp bewust herkent.

Een inverse vraag kan meerdere antwoorden hebben

Een sterk beschadigd beeld kan bij veel mogelijke originele beelden passen. Het model gebruikt patronen uit training en eventuele conditionering om een uitkomst te kiezen. Dat is bruikbaar voor generatie, maar bij restauratie of wetenschap moet je onderscheid maken tussen gemeten details en aangevulde structuur. Een helder geworden cijfer kan er overtuigend uitzien zonder dat de opname dat cijfer uniek bevatte. In een brongebonden reconstructie rapporteer je onzekerheid, de meetketen en de aannames; in een illustratie noem je het beeld als illustratie.

GRENS

Het panorama is een illustratie van een proces. Het is geen reeks werkelijk opgeslagen toestanden van een geteste diffusierun.



VECTOR • VAN SIGNAAL NAAR MODEL

Conditionering geeft de generatie richting

Tekst, beelden of andere invoer kunnen de uitkomst sturen. Guidance versterkt de koppeling, maar een hogere instelling is geen universele kwaliteitsknop.

Een tweede informatiestroom

Een tekstencoder zet de opdracht om in representaties. Het generatieve model gebruikt die naast de huidige beeldtoestand. In latent diffusion wordt de ruisprocedure uitgevoerd in een geleerde compacte beeldruimte en wordt het resultaat naar pixels gedecodeerd. Dat bespaart ruimtelijk rekenwerk, maar de encoder en decoder voegen hun eigen representatiekeuzes en verliezen toe. Een pixelbeeld en zijn latente code zijn niet simpelweg hetzelfde bestand met minder bytes.

Twee voorspellingen combineren

Classfier-free guidance combineert conditionele en onconditionele voorspellingen tijdens sampling. Een gebruikelijke schrijfwijze is ϵ

gestuurd = ϵ onvoorwaardelijk + $s \cdot (\epsilon$ voorwaardelijk - ϵ onvoorwaardelijk). Met $s = 1$ krijg je in die notatie de conditionele voorspelling. Grotere s versterkt het verschil. De exacte parameterconventie kan per implementatie verschillen; vermeld dus de gebruikte regel. De uitkomst kan beter bij de opdracht passen en tegelijk minder gevarieerd of met zichtbare artefacten worden.

Een seed is een afspraak over toeval

Een vast beginruispatroon helpt een verandering in opdracht of instelling onderzoeken. Houd model, sampler, stappen en resolutie gelijk als je één factor wilt vergelijken. Een seed alleen vertelt niet alles over reproduceerbaarheid bij andere hardware of implementaties. Beschrijf ook wat de tekst wel en niet vraagt: een redelijke generator kan een object toevoegen dat nergens verplicht was. Bij een illustratie is dat een ontwerpskeuze; bij een documentair bewijs zou het de interpretatie veranderen. Label nieuwe generatieve beelden en verwissel hun plausibiliteit niet met een waargenomen gebeurtenis. ♦



AXI Een sterke koppeling aan een opdracht vertelt nog niet dat de afgebeelde scène werkelijk is gebeurd.

Quiz: tijdvensters en relatieve schaal

Een sample, een frame en een meter zijn verschillende telobjecten. De aangenomen instellingen houden iedere som controleerbaar.

Twee miniaturen, twee eenheden

De audio-kaartjes op pagina 18 gebruiken 16 kHz en vensters van 400 samples; het schaalvoorbeeld staat op pagina 25. Uitwerkingen staan op pagina 49.

1 Een audiofragment van 2 s is opgenomen met 16.000 samples/s. Hoeveel samples bevat het?

- ☐ A 8000
- ☐ B 32.000
- ☐ C 16.000

Reken hier

2 Een venster van 400 samples bij 16 kHz duurt hoeveel milliseconden?

- ☐ A 25 ms
- ☐ B 40 ms
- ☐ C 2,5 ms

Reken hier

3 Een 1000-samplefragment krijgt vensters van 400 en stap 200, zonder padding. Hoeveel volledige vensters passen?

- ☐ A Vier
- ☐ B Drie
- ☐ C Vijf

Reken hier

4 Relatieve diepte $r=2$ hoort bij een referentie op 3 m. Het fictieve model gebruikt $Z=ar$. Wat is Z voor $r=5$?

- ☐ A 5 m
- ☐ B 7,5 m
- ☐ C 10 m

Reken hier

Jouw antwoorden

1 2 3 4

■ VECTOR · ONDERZOEK MET AANNAMEN

Een relatieve dieptekaart krijgt een maatlat

Een fictief model levert verhoudingen. Eén betrouwbare referentie kan de gekozen schaalvariant in meters omzetten.

ACHTERGROND

Voorspellen is een taak

DPT gebruikt transformerrepresentaties en een decoder voor dichte voorspellingen, waaronder monoculaire diepte. Training kan een bruikbare kaart leren. De betekenis van de output—relatief, inverse diepte of metrisch—moet uit de precieze modelvariant blijken; een kleurplaatje alleen specificeert die eenheid niet.

Bron: 33

De afspraak is $Z = ar$

Ons eigen model geeft $r=2$ voor de referentiekaart en $r=5$ voor een voorwerp. We nemen aan dat één positieve schaal a de volledige kaart vermenigvuldigt. Een gemeten referentiediepte van 3 m geeft $a=1,5$. De andere diepte wordt 7,5 m. Bij een model met extra onbekende offset zou één referentie niet genoeg zijn.

Meet waar de schaal vandaan komt

De referentie moet bij dezelfde camera-instelling en bij de gebruikte r -waarden horen. Noteer haar onzekerheid en controleer een tweede ongebruikte afstand. Anders kan een mooi omgerekende kaart verborgen schaalfouten houden. Het woord “meter” wordt pas gerechtvaardigd door die kalibratie-afpraak.

KORTE KIJKNOTITIE

Een kleurrijke legenda kan misleiden

Rood hoeft niet altijd dichtbij te betekenen.

Bewaar het label

Een kaart kan afstand, inverse afstand of alleen een genormaliseerde waarde tonen. Gebruik een legenda met de getoonde grootte en haar schaal. Vergelijk twee kleurenkaarten pas nadat je die omzetting hebt gecontroleerd.

Bronnen: 31, 33

Quiz: luisteren, genereren en herkomst

Zes vragen onderzoeken wat de procedure toevoegt en wat de meting werkelijk bevat. De juiste uitspraak maakt haar grens expliciet.

Volg het signaal

Voor sampling ligt een fysisch signaal. Daarna komen digitale waarden, representaties en voorspellingen. Bij generatie begint de procedure juist met een gekozen invoer of ruis. Een technische herkomstketen beschrijft een bestandsgeschiedenis, geen universele waarheid van de scène.

1 Een signaal bevat een cosinus van 10 kHz. Sampling gebeurt op 16 kHz zonder passend antialiasfilter. Welke frequentie kan dezelfde samplewaarden hebben?

- (A) 16 kHz.
- (B) 6 kHz.
- (C) Geen enkele andere frequentie is mogelijk.

Waarom?

2 Een transcript vult tijdens een onduidelijk fragment een plausibele naam in. Wat is de beste controle?

- (A) Luister naar het oorspronkelijke fragment en toets of de naam hoorbaar is.
- (B) Kies automatisch de langste transcriptvariant.
- (C) Beschouw correcte spelling als bewijs dat de naam werd gehoord.

Waarom?

3 Welk CTC-pad levert na herhalingsreductie en blankverwijdering twee letters L op?

- (A) L, L, blank.
- (B) blank, L, L.
- (C) L, blank, L.

Waarom?

4 Een diffusiegenerator begint met een ruissample. Welke uitleg is te sterk?

- (A) Het model gebruikt geleerde patronen voor de omgekeerde procedure.
- (B) Er ligt altijd één werkelijk origineel verstopt in iedere nieuwe ruis.
- (C) De sampler en ruisplanning beïnvloeden de uitkomst.

Waarom?

5 Latent diffusion voert de ruisprocedure uit in welke ruimte?

- (A) Een geleerde latente beeldrepresentatie, met een decoder naar pixels.
- (B) Altijd een exact kopieerbare lijst originele pixels zonder verlies.
- (C) Een ruimte die automatisch alle verborgen meetdetails bewaart.

Waarom?

6 Een geldige C2PA-keten zegt eerlijk dat een beeld is gegenereerd. Wat volgt daaruit?

- (A) De bergen op het beeld zijn noodzakelijk door een camera waargenomen.
- (B) Iedere ontbrekende credential op een ander beeld bewijst bedrog.
- (C) De herkomstclaim kan controleerbaar zijn terwijl de scène fictief is.

Waarom?

Jouw antwoorden

1	2	3
4	5	6

■ VECTOR · VAN SIGNAAL NAAR MODEL

Een GAN leert tegenover een beoordelaar

Een generator maakt samples; een discriminator probeert de trainingsdata van de gegenereerde samples te onderscheiden. Beide krijgen een gekoppeld trainingssignaal.

Een spel met twee netwerken

De generator zet een latente invoer, vaak ruis, om in bijvoorbeeld een beeld. De discriminator ontvangt echte trainingssamples en gegenereerde samples. Zijn doel helpt het systeem verschillen te vinden; het generatorverlies belooft uitkomsten die de discriminator minder goed onderscheidt. De woorden echt en nep slaan hier op herkomst uit de trainingsset of generator. Ze betekenen niet dat ieder trainingsbeeld een ware gebeurtenis documenteert.

Het signaal kan veranderen

Als de discriminator leert, verandert wat voor de generator moeilijk is. De optimalisatie van beide kanten is daardoor gekoppeld. Een veelbesproken probleem is mode collapse: de generator produceert een beperkt aantal patronen terwijl de gewenste verdeling rijker is. Zeer overtuigende voorbeelden kunnen dan naast slechte dekking bestaan. Laat bij een beoordeling daarom niet alleen de mooiste afbeelding zien. Onderzoek variatie, fouten en de manier waarop voorbeelden zijn geselecteerd.

Een scherp beeld is een eigen maatstaf

GANs kunnen visueel scherpe samples opleveren, maar scherpte beantwoordt niet alle vragen over overeenkomst, diversiteit of betrouwbaarheid. Een bedachte neus kan perfect scherp zijn en toch niet bij de gemeten persoon horen. Bij conditionele generatie moet je ook testen of de gegeven labels of broninformatie worden gevolgd. Het originele

Een scherp beeld is een eigen maatstaf

GAN-onderzoek beschrijft het adversariële trainingsidee; latere architecturen gebruiken aanvullende technieken. Houd het basismodel en een specifiek productsysteem uit elkaar. De volgende pagina kiest een andere route: een expliciete latente verdeling en reconstructie, met een eigen afweging.

WOORDEN IN HET TRAININGSSPEL

“Echt” betekent in het GAN-doel: afkomstig uit de aangeboden trainingsdata. Een fictieve illustratie kan ook een echt trainingssample zijn.

VECTOR · VAN SIGNAAL NAAR MODEL

Een VAE ordent een latente verdeling

Een variational autoencoder leert een verdeling van codes voor een invoer en een decoder voor de reconstructie. Het doel balanceert passen bij data met een geordende codewereld.

Een code is geen enkel vast punt

Een encoder schat bijvoorbeeld een gemiddelde en spreiding voor een latente verdeling. Tijdens training wordt een code gesampled en door de decoder verwerkt. De reconstructieterm beoordeelt hoe de uitkomst bij de invoer past volgens een opgegeven waarschijnlijkheidsmodel. Een aanvullende term stuurt de codeverdeling richting een gekozen prior. De precieze doeluitwerking komt uit de variationale formulering. Het is meer dan een gewoon compressieprogramma dat alleen pixels terug wil kopiëren.

Waarom regularisatie nodig is

Zonder samenhang in de codewereld kun je invoeren reconstrueren terwijl willekeurige codes weinig bruikbaar opleveren. Een prior maakt sampling en interpolatie beter hanteerbaar, maar kan spanning geven met fijne reconstructiedetails. Te sterke of ongunstige regularisatie kan latente informatie onvoldoende benutten. Het zichtbare resultaat hangt ook af van de decoder en verlieskeuze. Het populaire beeld van “VAEs zijn wazig” is daarom geen complete verklaring van de methode.

Vergelijk op de juiste taak

Een GAN gebruikt een discriminator als trainingssignaal; een VAE gebruikt een variationale opzet met reconstructie en een latente verdeling. Diffusie leert weer een ruisgerelateerde procedure. Alledrie kunnen onderdeel van complexere systemen zijn. Stel eerst de taak vast: nieuwe samples maken, een beeld reconstrueren, een code leren of een meting aanvullen. Kies daarna passende criteria. Een betere waarde op één criterium geeft geen universeel winnaaroordeel. Bij het herstellen van een meetbeeld blijft de centrale

Vergelijk op de juiste taak

vraag welke details door het oorspronkelijke signaal worden ondersteund en welke door het generatieve model worden gekozen.

CODEWERELD

Interpoleren met een grens

Een vloeiende overgang kan nuttig zijn zonder een natuurkundige route te volgen.

Tussen twee codes

Een rechte lijn tussen twee latente codes kan tussenbeelden geven. Zij beschrijft een gekozen geometrische route in het model, geen werkelijk proces waarmee het ene gefotografeerde object in het andere veranderde. Laat de codekeuze en de betekenis van de proef apart zien.

Bron: 16

Het venster en de schaalkaart

Drie puzzels combineren tijd en verhoudingen. Ze gebruiken alleen opgegeven eindige waarden; uitwerkingen staan op pagina 48.

1 Een reeks heeft samples 0 t/m 11. Vensters zijn vijf samples lang en starten iedere drie samples, zonder padding. Welke starts zijn toegestaan?

Mijn antwoord

2 Een fictief model geeft relatieve waarden $r_1=2$ en $r_2=6$. Je weet alleen dat $Z=ar$ met onbekende $a>0$. Welke diepteverhouding staat vast, en zijn meters uniek bepaald?

Mijn antwoord

3 Bij datzelfde model blijkt de eerste diepte exact 4 m. Welke a en tweede diepte volgen uniek?

Mijn antwoord

De randen tellen mee

Een venster moet volledig binnen het fragment vallen. Een schaalparameter werkt in alle punten van dezelfde fictieve modelvariant gelijk.

Ontwerp een tegenfoto voor het dieptemodel

Bouw op papier een proef met een platte kaart en een ruimtelijk voorwerp. Een schets van de meetopstelling is hier je werkproduct.

01 Maak de voorspelling controleerbaar

Teken een fotokaart van een auto naast een miniatuurauto. Noteer twee camerastanden en een bekende referentiemaat. Houd belichting en beeldgrootte zo gelijk mogelijk. Schrijf vooraf wat stereo moet zien: punten op de kaart horen bij één vlak; de miniatuur kan meerdere dieptes geven.

02 Bewaar ook een mislukte match

Kies een effen of spiegellend stukje dat geen betrouwbaar patroon geeft. Je voorspelling hoeft daar geen afstand te forceren. Geef dat gebied een apart symbool voor ongeldig. Noteer welke verandering extra informatie zou toevoegen, zoals een andere hoek of een herkenbaar patroon. Gebruik de tekening om de aannamen bespreekbaar te maken.

Mijn cameracontract

Teken camera links/rechts, baseline, platte kaart, miniatuur, referentiemaat en een gebied zonder betrouwbare match.

Vak 1

Drie ontbrekende onderscheidingen

*Elke proef geeft twee werelden die dezelfde zichtbare informatie kunnen opleveren.
Vraag welke aanvullende gegevens het verschil zouden kunnen tonen.*

Twee patronen, één grijs vak

Een speelgoedcamera geeft van twee invoerpixels alleen het gemiddelde door. patroon A heeft waarden $[0,100]$; patroon B heeft $[50,50]$. Beide worden één pixel met waarde 50. Je ziet uitsluitend die laatste waarde en kent de gemiddelde-regel.

Kun je uniek vaststellen welk patroon voor de camera lag?

Denk aan:

- Reken beide gemiddelden na.
- Welke extra meting zou de twee patronen onderscheiden?

Jouw keuze en redenering

Eén kanaal, twee bronnen

Een opnamekanaal geeft alleen $s(t) = a(t) + b(t)$. De som is overal nul. Mogelijkheid A: beide bronnen zijn nul. Mogelijkheid B: $a(t) = \sin(t)$ en $b(t) = -\sin(t)$. Geen andere meting of informatie is beschikbaar.

Bewijst de nulopname dat beide afzonderlijke bronnen stil waren?

Denk aan:

- Controleer de som in beide gevallen.
- Welke aanname maakt bronscheiding mogelijk of onmogelijk?

Jouw keuze en redenering

Een eerlijk ondertekende fictie

Een maker genereert een bergbeeld, registreert eerlijk het gebruikte maakproces en ondertekent de bijbehorende claim. De binding en handtekening valideren volgens het gekozen vertrouwensmodel. Er is geen opname van een camera op die berglocatie.

Welke twee uitspraken kun je wel en niet uit de validatie afleiden?

Denk aan:

- Wat is aan het bestand gebonden?
- Wat zou je nodig hebben om een werkelijk gebeurde scène te bewijzen?

Jouw keuze en redenering



Video vraagt samenhang door de tijd

Los mooie frames zijn nog geen overtuigende video. Voorwerpen, beweging, licht en camerastandpunt moeten over opeenvolgende beelden verenigbaar blijven.

Waarom zijn losse frames onvoldoende?

Een model dat ieder frame onafhankelijk genereert kan kleur, vorm of positie laten verspringen. Videomodellen gebruiken temporele informatie of andere gekoppelde representaties om gebeurtenissen tussen frames te verbinden. Video Diffusion Models onderzocht diffusie met ruimtelijke en temporele verwerking. De koppeling maakt het systeem beter geschikt voor een tijdreeks, maar bewijst geen exacte kennis van mechanica. Een voertuig kan nog een onmogelijke baan afleggen terwijl afzonderlijke beelden plausibel zijn.

Hoe toets je een beweging?

Volg bijvoorbeeld één gemarkeerde hoek van een doos door vijf frames. Is de route continu, blijft de doos dezelfde afmetingen houden en past een occlusie bij de overige zichtbare randen? Voor tekst op een draaiend voorwerp kun je controleren of letters dezelfde identiteit behouden. Zulke criteria zijn concreter dan alleen “het voelt echt”. Niet elke afwijking is een fout: een film kan een bewuste montage hebben. Leg daarom eerst vast welk type scène de test veronderstelt.

Welke rol speelt de camera?

Beweging in beeld kan door een bewegend object, een bewegende camera of beide ontstaan. Uit één korte zichtbare reeks is die ontleding soms niet uniek. Geluid en fysische aanwijzingen kunnen aanvullende informatie leveren. Bij generatieve video kan het model een plausibele combinatie kiezen zonder dat

Welke rol speelt de camera?

een werkelijke camera het heeft gemeten. Test tijdsconsistentie en brontrouw daarom afzonderlijk. De volgende pagina gaat uit van meerdere echte camerastandpunten en vraagt hoe je een nieuwe blik op dezelfde scène kunt berekenen.

+ KLEINE TEST

Kies één herkenbaar punt op een voorwerp en volg het door vijf frames. Noteer sprongen, occlusies en camerabeweging afzonderlijk. Verander het meetcriterium niet nadat je de beste uitkomst hebt gezien.



■ VECTOR · VAN SIGNAAL NAAR MODEL

Een radiantieveld berekent een nieuwe blik

NeRF en Gaussian splatting leren een scène uit meerdere beelden. Hun doel is een nieuwe cameraweergave, met andere aannames dan vrije videogeneratie.

NeRF verbindt positie en kijkrichting

Een Neural Radiance Field beschrijft kleur en dichtheid als functie van ruimtelijke positie en kijkrichting. Stralen van een virtuele camera worden bemonsterd en via volumetrische rendering samengevoegd tot pixels. De parameters worden aangepast zodat gerenderde standpunten bij de trainingsfoto's passen. De oorspronkelijke opzet gebruikt een statische scène en bekende of geschatte camerageometrie. Een object dat tussen opnamen verandert kan die aanname breken.

Wat achter de kast ligt

Als geen camera achter een kast heeft gekeken, bevat de dataset dat gebied niet rechtstreeks. Een representatie kan in een nieuw standpunt een plausibele invulling geven, maar die is niet noodzakelijk uniek uit de metingen bepaald. Bekijk prestaties daarom apart voor goed waargenomen gebieden en extrapolatie. Een spiegel of transparant object maakt de relatie tussen oppervlak en kleur bovendien complexer. Voor meetwerk wil je geometrische nauwkeurigheid toetsen; voor beeldweergave kan vooral visuele overeenkomst tellen. Een prachtige nieuwe blik en een betrouwbare maatvoering zijn twee verschillende resultaten.

Gaussische elementen geven een andere representatie

3D Gaussian splatting gebruikt een verzameling ruimtelijke Gaussische primitieve elementen met eigenschappen voor onder meer grootte, oriëntatie en verschijning. Een renderprocedure projecteert en mengt hun bijdragen. Dit is een expliciete representatie met een andere efficiëntie-afweging dan het oorspronkelijke impliciete NeRF-netwerk. Beide gebruiken observaties om nieuwe standpunten te benaderen. Het woord 3D betekent niet dat je automatisch een exacte, gesloten oppervlaktemesh krijgt.

ZO ZIT DAT

Drie vragen over dezelfde scène

Welke camerastandpunten zijn gebruikt? Welke gebieden waren daadwerkelijk zichtbaar? Welke foutmaat is voor nieuwe standpunten gemeten? Met die drie vragen kun je een afbeelding plaatsen zonder te doen alsof de hele verborgen wereld is gereconstrueerd.

Bronnen: 18, 19

Quiz: fouten tellen met ontbrekende pixels

Een systeem dat moeilijke punten overslaat kan zijn gemiddelde mooier laten lijken. Daarom krijgt de dekking een eigen getal.

De denominatoren zijn deel van het verslag

De achtpixelcasus op pagina 40 en de disparityrekening op pagina 13 zijn volledig aangenomen. Bekijk de uitwerkingen op pagina 49 na je eigen rekenspoor.

- 1** Voor vier geldige pixels zijn de absolute fouten 0,1; 0,2; 0,3; 0,2 m. Wat is hun gemiddelde?

- (A) 0,8 m
(B) 0,4 m
(C) 0,2 m

Reken hier

- 2** Een model geeft zes geldige uitkomsten voor acht referentiepunten. Wat is de dekking?

- (A) 75%
(B) 6%
(C) 80%

Reken hier

- 3** Een andere variant beoordeelt slechts twee makkelijke punten. Is haar kleinere gemiddelde fout automatisch beter voor de volledige taak?

- (A) Ja, de laagste fout wint altijd.
(B) Nee, geldige dekking en dezelfde taakpunten moeten meewegen.
(C) Ja, ontbrekende outputs zijn altijd correct.

Reken hier

- 4** Met $fb=80$ is bij $d=19$ de diepte ongeveer 4,2105 m en bij $d=21$ ongeveer 3,8095 m. Welke conclusie klopt?

- (A) De diepteband is exact symmetrisch rond 4 m.
(B) Een pixel is altijd precies 20 cm.
(C) De inverse omzetting geeft hier ongelijke afwijkingen.

Reken hier

Jouw antwoorden

1 2 3 4

De wereld die even groot lijkt

Een beeld kan goed voorspelde patronen bevatten en toch twee fysiek verschillende werelden toelaten.

Alles tweemaal zo groot

Een ideale pinholecamera fotografeert een object. Wereld B verdubbelt iedere objectlengte én iedere afstand tot de camera. Belichting en zichtbare oppervlaktetextuur zijn in deze gedachtewereld zo gekozen dat het beeld gelijk blijft.

Kan dit ene identieke beeld de absolute schaal tussen A en B vaststellen?

Denk aan:

- Welke verhouding blijft gelijk?
- Welke aparte meting zou schaalinformatie toevoegen?

Mijn voorspelling en waarom

Twee blikken op twee tijden

Twee camera's maken dezelfde objectfoto, maar camera rechts neemt 0,1 s later op. Het object beweegt ondertussen. Een algoritme behandelt de opnames als gelijktijdig.

Kan de gemeten verschuiving uitsluitend door camerageometrie worden verklaard?

Denk aan:

- Welke tweede oorzaak krijgt hetzelfde zichtbare spoor?
- Wat hoort in het opnamecontract?

Mijn voorspelling en waarom

Rekenwerk: patches, samples en scores

Vier berekeningen laten zien waarom een eenheid en een meetafpraak bij een getal horen. Reken eerst zelf, kies daarna de beste uitkomst.

Vier expliciete afspraken

We tellen alleen beeldpatches, gebruiken onbewerkte zestienbitmono-audio, nemen achtbitgrijswaarden met maximum 255 en vergelijken twee opgegeven vectoren. Compressie, extra tokens en systeemoverhead vallen buiten deze vragen. De volledige tussenstappen staan op pagina 47.

- 1 Een beeld van 448×448 gebruikt patches van 16×16 . Hoeveel beeldpatches krijg je, zonder extra tokens?**

- (A) 392.
(B) 784.
(C) 196.

Reken hier

- 2 Hoeveel onbewerkte audiobytes bevat één minuut mono bij 16.000 samples per seconde en twee bytes per sample, zonder header?**

- (A) 1.920.000 bytes.
(B) 960.000 bytes.
(C) 32.000 bytes.

Reken hier

- 3 Achtbitgrijswaarden hebben maximum 255. $MSE = 100$. Wat is $10 \log_{10}(255^2/100)$?**

- (A) 48,13 dB.
(B) 6,50 dB.
(C) Ongeveer 28,13 dB.

Reken hier

- 4 Welke cosinusovereenkomst hebben [3,4] en [4,3]?**

- (A) 0,48.
(B) 0,96.
(C) 1,00.

Reken hier

Jouw antwoorden

1 2 3 4

■ VECTOR · VAN SIGNAAL NAAR MODEL

Textuur kan de verkeerde aanwijzing zijn

Een classifier kan een patroon gebruiken dat vaak samen met het label voorkwam. Een gecontroleerde verandering laat zien hoe robuust die aanwijzing is.

Wat gebeurt er als vorm en textuur botsen?

Geirhos en collega's maakten onder meer beelden waarin vorm en textuur naar verschillende objectcategorieën wijzen. Onderzochte ImageNet-getrainde CNNs bleken vaak sterker textuur te volgen dan mensen in de vergelijking. Training met anders gestileerde beelden kon het gebruik van vorm beïnvloeden. Het resultaat is een gemeten gedrag in een specifieke opzet. Het betekent niet dat ieder CNN uitsluitend textuur gebruikt of dat alle menselijke herkenning altijd vorm volgt.

Hoe verschilt een aanval van gewone ruis?

Adversarial examples zijn invoeren die doelgericht worden aangepast om een model tot een andere uitkomst te brengen. Een verandering kan voor mensen klein zijn en voor de beslisgrens van het model groot. Dat is een andere vraag dan gewone ruis toevoegen. De toegestane verstoring, kennis van het model en testdata bepalen de aanval. Een succesvolle aanval op één beeld bewijst geen universele onbruikbaarheid, maar toont wel een concrete kwetsbaarheid van die opstelling.

Volgt het label het bedoelde object?

Als je mokken wilt herkennen, moeten kleur of achtergrond niet de enige route naar het label zijn. Maak een test met dezelfde voorwerpen op meerdere achtergronden en met verschillende belichting. Voeg ook andere objecten op dezelfde achtergrond toe. Zo scheid je objectinformatie van meereizende context. Een model kan tegelijk gemiddeld goed scoren en één systematische shortcut gebruiken. De uitsplitsing is daarom informatiever dan een enkel totaalpercentage. Noteer welke veranderde beelden nog geldig bij dezelfde taak horen; een volledig verborgen object is geen eerlijke herkenningsproef meer.

TOETSVRAAG

Blijft het bedoelde voorwerp herkenbaar nadat je een factor verandert? Als de informatie zelf verdwijnt, meet je een andere taak dan robuust herkennen.

VECTOR • VAN SIGNAAL NAAR MODEL

De sensor beslist wat ontbreekt

Meer contrast of slimme nabewerking kan een beeld leesbaarder maken. Het kan niet altijd uniek terughalen wat de meting niet onderscheidde.



FIG. 1 Foto • Cosmic Cliffs: astronomische waarnemingen met toegewezen kleuren. Het beeld dient als voorbeeld van meetgegevens en visualisatiekeuzes.

1. Gemeten signaal

Detectorbanden leveren gegevens; een beeldmodel is niet de bron van die meting.

2. Weergavekeuze

Toegewezen kleur helpt patronen zien en vraagt uitleg in het bijschrift.

Fotonen komen niet exact gelijk aan

Bij een eenvoudige Poissonopzet is de standaardafwijking van een telling N gemiddeld ongeveer $\sqrt{N}$. Bij gemiddeld 100 gedetecteerde fotonen is de spreiding ongeveer 10 en de relatieve spreiding 10 procent. Bij 10.000 is zij ongeveer 100, relatief 1 procent. Dit speelgoedvoorbeeld laat alleen shot noise zien. Een echte camera heeft ook achtergrond, uitleesruis, kwantumefficiëntie en andere effecten. Een hogere versterking maakt niet vanzelf extra fotonen.

Resolutie is geen volledig detailbewijs

Een lens vervaagt fijne structuren; pixels integreren licht over een oppervlak en tijd. Als twee verschillende patronen na die keten dezelfde

pixelwaarden geven, is reconstructie uit die waarden alleen niet uniek. Een model kan op basis van training een scherp patroon kiezen. Dat kan bruikbaar zijn als illustratieve verbetering, maar het gekozen detail is dan mede een prior. Dezelfde grens geldt bij een onleesbaar teken in een foto: één scherp resultaat bewijst niet dat de letter zo op het origineel stond.

De kosmische foto vraagt een bijschrift

De nevelafbeelding in dit nummer toont een verwerking van astronomische waarnemingen. Toegewezen kleuren maken gemeten banden zichtbaar; zij zijn niet noodzakelijk de kleuren die een oog op die plaats zou zien. Dat is een expliciete visualisatieafpraak, geen fout. Houd haar apart van generatief ingevulde details. Wie wetenschappelijke beelden beoordeelt, zoekt de detector, banden, verwerking en onzekerheid op. Een foto kan echte gegevens bevatten én keuzes in representatie. Het bijschrift maakt die combinatie beter leesbaar. ♦

DATAPUNT



standaardafwijking in het ideale Poissonvoorbeeld

■ VECTOR · VAN SIGNAAL NAAR MODEL

Een beeldscore heeft een meettaak

Pixelverschil, structurele overeenkomst en verdelingsafstand beoordelen andere eigenschappen. Geen enkele waarde vat alle beeldkwaliteit samen.

Drie verschillende controles

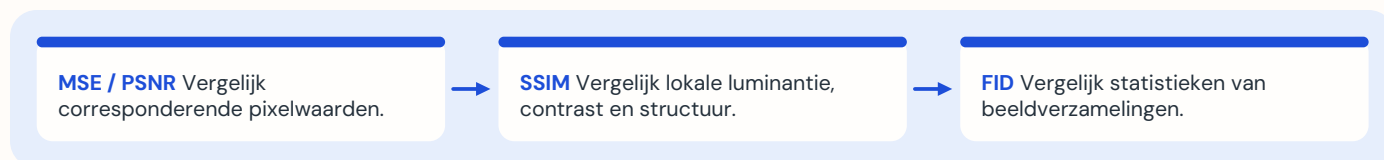


FIG. 1 Deze kaartjes zijn definities, geen gemeten ranglijst.

Een exact pixelcriterium

Mean squared error is het gemiddelde van gekwadrateerde verschillen tussen corresponderende waarden. Bij achtbitgrijswaarden en $MSE = 100$ is $PSNR = 10 \log_{10}(255^2/100)$, ongeveer 28,13 dB. De maximumwaarde 255 is een expliciete conventie. Een kleine verschuiving van een overigens gelijk beeld kan veel pixelverschil opleveren. De maat meet uitlijning en fouten in die getallen; zij beslist niet rechtstreeks of een afbeelding overtuigend of feitelijk correct is.

Lokale structuur vergelijken

SSIM vergelijkt aspecten van luminantie, contrast en structuur in lokale vensters. Dat sluit bij sommige verschillen beter aan op visuele beoordeling dan een kale kwadratische fout. De vensters, stabiliserende constanten en implementatie beïnvloeden de uitkomst. Een hoge structurele overeenkomst kan tegelijk een belangrijk klein detail missen. Controleer bij een medische tekst of een getal daarom niet alleen een algemene beeldmaat, maar het specifieke taakdetail.

Een verzameling samples beoordelen

FID vergelijkt statistieken van representaties van twee beeldverzamelingen, gewoonlijk met een vast featuremodel. Het is geen score voor de waarheid van één afbeelding. Het aantal samples, voorbewerking en representaties horen bij de meetopstelling. Een generator kan een betere gemiddelde verdelingsafstand halen en toch

ongewenste memorisatie of rare tekst produceren. Kies dus meerdere criteria: visuele variatie, opdrachttrouw, fouten op relevante details en eventueel herkenbaarheid. Rapporteer de scores met hun betekenis. Zonder die betekenis kan een kleine numerieke winst een groot maar onterecht verhaal krijgen. ♦

DATAPUNT

28,13 dB

PSNR in het opgegeven achtbitvoorbeeld

Claude Opus 5.5: een beeld lezen en pixels overslaan

Peildatum 8 oktober 2026. Claude Opus 5.5 verwerkt beelden, maar productcapaciteit en de volledigheid van een meettaak krijgen ieder hun eigen toets.

Wat de documentatie en de proef onderscheiden

claude-opus-5-5 verscheen volgens Anthropic op 22 september 2026 en heeft tekst/beeld-invoer, tekstuitvoer en 1M context. Anthropic meldt voor Terminal-Bench 4.0 xhigh-effort bij Opus 5.5, en voor GPT-6 Astra high; de cijfers zijn dus niet één uniforme instellingenproef. Hun Opus 5-reproductie is 52,3% naast 51,8% op de publieke lijst, volgens de voetnoot binnen de ruis. Zo'n codebenchmark meet geen disparity of dieptedekking. De volgende achtpixelcasus is daarom expliciet fictief; wij hebben deze modellen niet op de punten uitgevoerd.

Variant A

A rapporteert voor alle acht punten fouten van 0,1; 0,2; 0,3; 0,2; 1,0; 1,2; 1,5; 1,5 m. De totale absolute fout is 6,0 m; het gemiddelde is 0,75 m. De geldige dekking is $8/8 = 100\%$. Dit gemiddelde zegt niets over de verdeling zonder de afzonderlijke waarden.

Variant B

B meldt slechts de eerste vier punten als geldig, met dezelfde fouten 0,1;0,2;0,3;0,2. Daar is het gemiddelde 0,20 m en de dekking $4/8 = 50\%$. De vier moeilijkere punten zijn ontbrekende outputs, geen nul-fouten. Alleen 0,20 naast 0,75 zetten belooft selectie.

Een complete beoordelingsregel

Bewaar de fout op gezamenlijk geldige punten, de volledige dekking en de plekken waar outputs ontbreken. Een echte toepassing kan expliciet toestaan dat onzekere punten aan een andere route worden doorgegeven. Schrijf dan die taak en kosten vooraf op. Een derde fictieve variant met zes geldige punten heeft dekking 75%; dat percentage is hier een outputaandeel, geen kansinterval. Hetzelfde woord “dekking” vraagt dus een definitie.

■ VECTOR · VAN SIGNAAL NAAR MODEL

Herkomst is een controleerbare keten

Content Credentials kunnen claims over ontstaan en bewerkingen aan een mediabestand verbinden. De handtekening en de bewering moeten allebei worden gelezen.

Wat maakt de handtekening controleerbaar?

Een digitaal manifest kan herkomstclaims bevatten en cryptografisch aan een bestand worden gebonden. Validatie controleert onder meer die binding en de handtekening volgens een vertrouwensmodel voor de ondertekenaar. Dat helpt wijziging of verwisseling van gegevens signaleren. De C2PA-specificatie beschrijft hiervoor een technische structuur. Een geldige claim kan ook melden dat een beeld met een generator is gemaakt. Juist die duidelijke vermelding maakt de oorsprong informatiever.

Welke vraag blijft open?

Een geldige handtekening maakt niet iedere inhoudelijke uitspraak in het beeld waar. Een camera kan een geënceneerde situatie vastleggen; een generator kan een fictieve scène eerlijk als gegenereerd registreren. Ook de betrouwbaarheid van een ondertekenaar is een afweging binnen een vertrouwensmodel. Lees daarom welke gebeurtenis, tool of bewerking is vastgelegd. Een groen vinkje zonder zicht op de claim laat veel informatie weg. De standaard geeft geen universeel waarheidslabel voor de afgebeelde wereld.

Wat betekent een ontbrekende keten?

Afwezigheid van credentials bewijst evenmin dat een beeld nep is. Metadata kan ontbreken, verwijderd worden of nooit zijn toegevoegd. Aanwezigheid en afwezigheid vragen dus een verschillende interpretatie. Voor eigen onderzoek bewaar je bronbestand, opgegeven maakproces, versie en eventuele

Wat betekent een ontbrekende keten?

validatie-uitkomst. Controleer daarnaast de feitelijke bewering met onafhankelijke informatie. Het ene spoor gaat over bestand en geschiedenis, het andere over wat je ermee wilt bewijzen. Samen zijn ze sterker dan een los zegel, zonder dat ze alle onzekerheid laten verdwijnen.

LEES DE CLAIM

Ondertekenaar, binding, geregistreerde bewerking en afgebeelde gebeurtenis zijn vier verschillende gegevens. Een technisch geldige keten mag je nauwkeurig benoemen zonder haar als algemene waarheidsgarantie te presenteren.

VECTOR · VAN SIGNAAL NAAR MODEL

Een dataset kan groepen ongelijk tonen

Gemiddelde nauwkeurigheid kan verschillen tussen opnameomstandigheden verbergen. Beschrijf zowel de verzameling als de fouten in duidelijke deelgroepen.

Rapporteer naast het totaal



FIG. 1 Een onderzoeksvolgorde voor de fictieve groepen.

Een gemiddelde kan weinig vertellen

Een fictieve test heeft negentig heldere foto's met 81 juiste labels en tien donkere foto's met vier juiste labels. Het totaal is 85 van 100: 85 procent. De heldere groep haalt 90 procent, de donkere 40 procent. Het totaalcijfer heeft de slechte kleinere groep niet weerlegd; het heeft haar minder gewicht gegeven. Deze cijfers zijn een didactische dataset, geen gemeten prestatie van een werkelijk model.

Documenteer de verzameling

Datasheets for Datasets stelt vragen over motivatie, verzameling, samenstelling, verwerking en gebruik. Model Cards legt onder meer beoogd gebruik, evaluatie en verschillen tussen groepen vast. Documentatie vervangt geen test, maar maakt een test beter te interpreteren. Bij beelden kunnen camera, belichting, achtergrond, geografische herkomst en labelprocedure samen verschillen. Bij spraak tellen taal, stem, opnameapparatuur en ruis mee. Een groepsverschil vraagt vervolgonderzoek naar die verbonden factoren.

Een kleine groep vraagt voorzichtigheid

Tien voorbeelden geven weinig zekerheid over een breed domein. Verzamel meer vooraf gekozen gevallen en rapporteer aantallen naast

percentages. Laat ontbrekende categorieën zichtbaar in plaats van ze stilzwijgend onder een gemiddeld oordeel te schuiven. Een model aanpassen om één groep te verbeteren kan andere groepen veranderen; toets de nieuwe versie opnieuw. De nuttige uitkomst is een concrete kaart van waar de huidige opzet werkt, faalt of nog onvoldoende is onderzocht. Dat helpt kiezen of het systeem bij de bedoelde taak past. ♦

DATAPUNT

85%

totaal in de fictieve test

DATAPUNT

40%

juistheid binnen de tien donkere foto's

■ VECTOR · VAN SIGNAAL NAAR MODEL

De werkplaats: vier voorwerpen, één factor

Met eigen foto's kun je een kleine herkenningstest ontwerpen. De camera-opnamen zijn de gecontroleerde input; het modelantwoord is de te beoordelen uitkomst.



FIG. 1 Foto • Een bos toont hoe achtergrond en overlappende vormen visuele herkenning complexer kunnen maken; de voorwerpenproef gebruikt eigen opnamen.

Zo doe je het

- 1** Kies vier gewone voorwerpen zonder persoonlijke gegevens.
- 2** Fotografeer ieder op een rustige achtergrond.
- 3** Schrijf vooraf labels, onbekendcategorie en beoordelingsregels op.
- 4** Bewaar systeemversie, vraag, instellingen en ieder antwoord.
- 5** Verander alleen de achtergrond voor de tweede ronde.
- 6** Verander belichting in een aparte derde ronde.
- 7** Rapporteer aantallen, concrete fouten en de grens van de kleine test.

+ LOGBOEK

Voorwerp: ... Achtergrond: ... Licht: ... Antwoord: ...
 Juist label: ... Toegevoegd detail: ... Veranderde factor: ...

01 Begin met een herkenbare taak

Kies een mok, sleutel, lepel en boek. Fotografeer elk op een eenvoudige achtergrond bij vergelijkbare belichting. Gebruik geen gezichten, persoonlijke documenten of herkenbare adressen. Vraag een beschikbaar beeldsysteem naar één voorwerp per foto, of vergelijk eerder verzamelde antwoorden op papier. Leg vooraf vast welke labels juist zijn en wat telt als onbekend. De proef vereist geen ingebouwde AI-dienst in dit magazine.

02 Wissel de achtergrond, behoud de vorm

Maak daarna nieuwe opnamen met een dikkere achtergrond. Houd afstand, licht en voorwerp zoveel mogelijk gelijk. In een aparte ronde verander je de belichting; meng beide veranderingen niet meteen. Noteer of het voorwerp voor jou nog herkenbaar is. Een totaal verborgen sleutel is geen test van dezelfde informatie. Bewaar de originele foto en eventuele crop, zodat de invoer achteraf controleerbaar blijft.

03 Beschrijf de fouten concreet

Schrijf bijvoorbeeld: vier van vier labels goed op de rustige achtergrond, twee van vier op de drukke, met een mok als vaas en een sleutel als mes. Dat is een bruikbaar klein resultaat, geen algemene score voor alle voorwerpen. Herhaal met nieuwe exemplaren en vooraf gekozen criteria. Kijk ook naar toegevoegde details: een antwoord kan het hoofdobject juist noemen en een niet-zichtbare inscriptie verzinnen. Bewaar de foutvoorbeelden naast de goede voorbeelden. Ze vertellen welke informatie de test echt heeft onderzocht en welke uitbreiding nog nodig is.

VECTOR • VAN SIGNAAL NAAR MODEL

Bouwstenen van kijken, horen en maken

Deze tijdlijn kiest een paar oorspronkelijke onderzoeken. De jaren markeren publicaties, geen claim dat alle bijbehorende ideeën toen voor het eerst bestonden.

1998 • Gedeelde filters voor tekens

LeCun en collega's beschreven gradient-based learning voor documentherkenning. Convolutionele gewichten delen lokale bewerkingen over een beeld. De taak was concreet: onder meer geschreven tekens herkennen, met een systeem waarin architectuur en training samenwerkten.

2006 • Labels zonder exacte tijdgrenzen

CTC bood een trainingsdoel voor sequenties waarvan de invoer langer is dan de gewenste labelreeks. De methode telt mogelijke alignments samen. Zij werd een belangrijke bouwsteen voor herkenning, naast andere encoder-decoder- en alignmentopzetten.

2012–2015 • Diepere beeldnetwerken

AlexNet liet sterke ImageNet-classificatie met een diep convolutioneel netwerk zien. ResNet onderzocht daarna residuele verbindingen om veel grotere diepte bruikbaar te optimaliseren. Een benchmarkwinst zegt iets over die taak en vergelijking, niet dat ieder visueel probleem is opgelost.

2013–2014 • Twee generatieve routes

Auto-Encoding Variational Bayes beschreef een variationale latentecode-opzet. Generative Adversarial Networks introduceerde een generator en discriminator met gekoppelde doelen. Hun

2013–2014 • Twee generatieve routes

verschillende trainingssignalen geven andere afwegingen in reconstructie, sampling en dekking van patronen.

2020 • Patches, radiantie en diffusie

Vision Transformer verwerkte een beeld als patchreeks. NeRF gebruikte een geleerd radiantieveld voor nieuwe camerastandpunten. DDPM onderzocht generatie via een omgekeerde ruisprocedure. Hetzelfde jaar leverde dus verschillende antwoorden op verschillende beeldvragen.

2021–2022 • Taal ontmoet media

CLIP leerde beeld-tekstvergelijking uit natuurlijke taal. Latent diffusion voerde een generatieve procedure in een compacte representatie uit. Whisper onderzocht grootschalige spraakherkenning met meerdere talen en taken. De gedeelde ambitie was bruikbare representaties; de meetdoelen bleven verschillend.

2023 • Nieuwe blikken en afbakening

3D Gaussian splatting onderzocht een expliciete scèneweergave met efficiënte rendering. Segment Anything onderzocht promptbare segmentatie. Een nieuwe blik, een masker en een objectnaam zijn elk afzonderlijke resultaten. De geschiedenis wordt duidelijker als je die vragen apart blijft benoemen.

Drie raadsels over dezelfde gegevens

Een label, een gemiddelde en een beeldvergelijking kunnen informatie verliezen. De regels hieronder maken precies zichtbaar welke conclusie wel volgt.

1 Een speelgoedcamera rapporteert het gemiddelde van twee waarden uit {0,50,100}. Het gerapporteerde gemiddelde is 50. Welke geordende invoerparen zijn mogelijk? Is de originele volgorde uniek vastgelegd?

Mijn antwoord

2 Een classifier mag alleen de labels “kat” en “hond” geven. Je biedt een foto van een mok aan. Het systeem kiest “kat” omdat die score hoger is. Bewijst dat dat de foto een kat bevat?

Mijn antwoord

3 Een reeks heeft vier binaire pixels, elk 0 of 1. Je weet alleen dat er precies twee enen zijn. Hoeveel verschillende reeksen blijven mogelijk?

Mijn antwoord

Zoek de overgebleven mogelijkheden

Voor ieder raadsel kun je een klein lijstje maken. Als meerdere gevallen dezelfde uitkomst geven, kun je

zonder aanvullende gegevens geen uniek geval aanwijzen. Alle oplossingen staan op pagina 47, inclusief de verleidelijke verkeerde stap.

Lees per paper de juiste vraag

Kies één resultaat dat je wilt begrijpen. Zoek daarna methode, data en foutmaat voordat je de conclusie op een nieuwe toepassing plakt.

01 Vision Transformer en CLIP • Representatie

Zoek hoe patches worden gemaakt en welke extra tokens een model gebruikt. Zoek bij CLIP hoe passende beeld-tekstparen tegenover andere paren worden gezet. De eerste paper onderzoekt beeldrepresentatie voor classificatie; de tweede verbindt beelden en taal. Noteer welke kandidaten een evaluatie aanbiedt en wat buiten die lijst valt.

02 Whisper en CTC • Signaal naar labels

Vergelijk de supervisie en de rol van tijdalignment. Een encoder-decoder voorspelt een tekstreeks; CTC telt passende tijdpaden op. Lees bij wav2vec 2.0 welk doel vóór gelabelde aanpassing wordt gebruikt. Het woord spraakmodel omvat dus verschillende trainingsroutes, niet één uniforme meetopstelling.

03 DDPM en latent diffusion • Nieuwe samples

Zoek de ruisplanning, voorspelde grootte en samplingprocedure. Kijk vervolgens hoe een encoder en decoder de latente ruimte definiëren. Verbeterde of variationale diffusieopzetten veranderen onderdelen van dezelfde familie. Een duidelijk verslag laat die keuzes zien, in plaats van alle modellen als één identiek algoritme te behandelen.

04 NeRF en Gaussian splatting • Standpunten

Zoek welke camera-informatie en scène-aannames zijn gebruikt. Bekijk de representatie en de test op nieuwe standpunten. Is een gebied werkelijk waargenomen of wordt er geëxtrapoleerd? Een visueel goede render zegt nog niet dat de geometrie overal exact gemeten is.

05 Objecthallucinatie en datasetdocumentatie • Fouten

Zoek hoe een extra object in een bijschrift wordt geteld. Vergelijk die vraag met de documentatie van datasets en modelkaarten. Een score kan een specifiek fouttype signaleren; een beschrijving van herkomst en groepen helpt plaatsen waar de score vandaan komt.

06 C2PA • Bestandsgechiedenis

Lees de scope en het vertrouwensmodel van de technische specificatie. Zoek welke binding, handtekening en claims worden gevalideerd. De standaard schrijft geen algemeen oordeel voor over de waarheid van de scène. Vertaal dat verschil in je eigen verslag naar twee vragen: hoe ontstond het bestand, en welke gebeurtenis bewijst het?

■ VECTOR • VAN SIGNAAL NAAR MODEL

Alle antwoorden, met berekeningen

Een juiste letter is het begin. Controleer de redenering tegen de aangegeven taak, informatie en eenheden.

Quiz • pagina 10

- 1 A. Rechts min links geeft $0-0 = 0$, $10-0 = 10$ en $10-10 = 0$. Een respons is nog geen objectnaam (pagina 4).
- 2 C. $y = x + F(x)$ voegt een geleerde verandering aan de korte invoerroute toe; prestaties blijven een toetsvraag (pagina 5).
- 3 B. $224/16 = 14$ en $14^2 = 196$ beeldpatches, zonder extra tokens (pagina 6).
- 4 A. CLIP vergelijkt beeld- en tekstvectoren die gezamenlijk contrastief worden getraind (pagina 7).
- 5 C. Het bijschrift voegt een niet-ondersteund object toe: objecthallucinatie (pagina 20).
- 6 B. Niet-waargenomen gebieden kunnen plausibel worden ingevuld zonder unieke meetsteun (pagina 33).

Quiz • pagina 26

- 1 B. 10 kHz en 6 kHz kunnen bij 16 kHz dezelfde cosinussamples geven: aliasing (pagina 14).
- 2 A. Luister naar de opname; een correct gespelde naam kan door taalvoorspelling zijn aangevuld (pagina 15).
- 3 C. L, blank, L blijft twee L-voorkomens na herhalingsreductie. Verwijder daarna de blank: LL (pagina 16).
- 4 B. Een nieuwe ruissample hoeft geen uniek bestaand

origineel te bevatten.

Sampling gebruikt geleerde patronen (pagina 22).

- 5 A. De ruisprocedure werkt in een geleerde latente representatie, daarna volgt decoding naar pixels (pagina 23).
- 6 C. Een valide claim kan eerlijk generatie melden. De scène hoeft daarom niet door een camera te zijn waargenomen (pagina 19, 41).

Rekenwerk • pagina 36

- 1 B. Langs elke rand passen $448/16 = 28$ patches. $28^2 = 784$, zonder extra tokens (pagina 6).
- 2 A. $16.000 \text{ samples/s} \times 2 \text{ bytes/sample} \times 60 \text{ s} = 1.920.000 \text{ bytes}$. Dat is 1,92 MB in decimale notatie, zonder header (pagina 11, 14).
- 3 C. $255^2 = 65.025$; delen door MSE 100 geeft 650,25. $10 \log_{10}(650,25) \approx 28,1308 \text{ dB}$ (pagina 39).
- 4 B. Het inproduct is $3 \times 4 + 4 \times 3 = 24$. Beide lengten zijn $\sqrt{(9+16)} = 5$. $\text{Cosinus} = 24/25 = 0,96$, geen waarheidskans (pagina 7).

Gedachtenproeven • pagina 31

Twee patronen: [0,100] en [50,50] leveren allebei gemiddeld 50. De enkele waarde identificeert de invoer niet uniek; extra metingen kunnen verschil maken.

Eén kanaal: nul + nul en $\sin(t) + (-\sin(t))$ geven beide nul. De afzonderlijke bronnen zijn uit alleen die som niet vast te stellen.

Ondertekende fictie: de gevalideerde herkomstclaim hoort bij het bestand binnen het gekozen vertrouwensmodel. Zij bewijst geen camerawaarneming van de bergscène, en meldt hier juist generatie.

Raadsels • pagina 45

1. [0,100], [50,50] en [100,0]; de volgorde is niet uniek. De som moet 100 zijn. Het gemiddelde laat de individuele waarden en volgorde open.

2. **Nee:** het bewijst alleen welke van de twee aangeboden kandidaten hoger scoorde. De mokcategorie ontbreekt. De keuze is geen bewijs dat het beeld een kat bevat.

3. **Zes:** 0011,0101,0110,1001,1010,1100. De combinatieformule $4 \times 3/2$ telt ieder gekozen paar maar één keer.

Simulatie • pagina 21

De amplitude is $2A \cdot |\cos(\varphi/2)|$. Bij $A = 1,5$: $\varphi = 0$ geeft 3; $\varphi = \pi/2$ geeft $1,5\sqrt{2} \approx 2,1213$; $\varphi = \pi$ geeft 0. Instelling 3,14 benadert π en laat ongeveer 0,00239 restamplitude. Veranderen van f verhoogt het aantal perioden op de dimensieloze as, niet deze amplitude. De opzet omvat geen demping, verschillende frequenties of volledige energiebalans.

Nieuwe puzzels: het volledige bewijs

De voorbeelden zijn aangenomen miniwerelden. Iedere oplossing gebruikt precies de aangegeven voorwaarden.

Raadsels • pagina 9

1. $d = 20$ pixels en $Z = 4$ m. De unieke patroonmatch geeft $d=42-22=20$. De diepte is $80/20=4$ m. Zonder de unieke overeenkomst zou de coördinatensom geen geldig meetbewijs zijn.

2. Nee. Er zijn twee toegestane matches. Gelijke patroonlabels geven hier geen extra onderscheid. Beide posities voldoen. Je moet aanvullende context gebruiken of de match als dubbelzinnig bewaren; één kiezen is geen logische afleiding.

3. 30, 20 en 10 pixels. Uit $d=fb/Z$ volgen $60/2=30$, $60/3=20$ en $60/6=10$. De gegeven verschillende dieptes leiden onder dit model tot verschillende disparities.

Raadsels • pagina 29

1. 0, 3 en 6. Start 0 eindigt op 4; start 3 op 7; start 6 op 10. Start 9 zou op 13 eindigen, buiten sample 11. Er zijn dus precies drie volledige vensters.

2. $Z2/Z1 = 3$; meters zijn niet uniek bepaald. De schaal valt uit de verhouding: $(6a)/(2a)=3$. $a=1$ geeft 2 en 6 m; $a=2$ geeft 4 en 12 m. Beide voldoen, zodat absolute meters extra informatie vragen.

3. $a = 2$ m per relatieve eenheid; $Z2 = 12$ m. $4=2a$, dus $a=2$. Vervolgens $Z2=6 \times 2=12$. De ene foutloze schaalreferentie bepaalt de ene onbekende schaalparameter in deze afspraak.

Gedachtenproeven • pagina 35

Alles tweemaal zo groot: Nee. Projectie gebruikt verhoudingen zoals objecthoogte gedeeld door

afstand; beide verdubbelen, zodat de verhouding gelijk blijft. Een bekende maat, gekalibreerde stereo-baseline of andere afstandsreferentie kan extra informatie geven. Dit sluit bruikbare voorkennis uit het voorbeeld. Een echt geleerd model kan schaal aannemen uit bekende objecten, maar die aanname is geen nieuw gemeten onderscheid tussen identieke beelden.

Twee blikken op twee tijden: Nee. Beweging kan naast de baseline beeldverschuiving veroorzaken. Leg synchronisatie vast of modelleer beweging met extra gegevens. De simpele static-scene-dieptesom is dan niet zonder meer geldig. Niet iedere tijdsafwijking maakt iedere match fout. Het punt is dat de beschreven geometrische aanname bij beweging apart moet worden getoetst.

■ VECTOR • BEWIJS BIJ DE NIEUWE OPGAVEN

Nieuwe quizzen: berekening en reden

De voorbeelden zijn aangenomen miniwerelden. Iedere oplossing gebruikt precies de aangegeven voorwaarden.

Quiz • pagina 8

1. A. $Z = fb/d = 800 \times 0,10 / 20 = 4$ m. Zie pagina 12.
2. C. $80/10 = 8$ m. Halve disparity geeft hier dubbele diepte. Zie pagina 12.
3. B. $d = 800 \times 0,20 / 4 = 40$ pixels. Zie pagina 12.
4. C. Correspondentie is niet uniek bepaald door de gegeven textuur. Een geldigheidsmasker bewaart dat gebrek aan informatie. Zie pagina 12.

Quiz • pagina 24

1. B. $2 \times 16.000 = 32.000$ samples. Zie pagina 18.
2. A. $400/16.000 \text{ s} = 0,025 \text{ s} = 25$ ms. Zie pagina 18.
3. A. De vensterstarts zijn 0,200, 400, 600 en de einden 399, 599, 799, 999: vier volledige vensters. Zie pagina 18.
4. B. $a = 3/2 = 1,5$ m per relatieve eenheid; $Z = 1,5 \times 5 = 7,5$ m. Zie pagina 25.

Quiz • pagina 34

1. C. De absolute fouten hebben som 0,8 m; hun gemiddelde is $0,8/4 = 0,2$ m. Zie pagina 40.
2. A. $6/8 = 0,75$. Dit is het aandeel punten met geldige output, niet intervaldekking. Zie pagina 40.
3. B. Het gemiddelde over een andere selectie vergelijkt niet dezelfde taak. Rapporteer ook ontbrekende punten. Zie pagina 40.
4. C. +0,2105 en -0,1905 m verschillen. Zie pagina 13.

Colofon • VECTOR 03

Dit nummer volgt de verbinding tussen fysieke signalen, geleerde representaties en nieuwe media.

Redactionele verantwoording

De tekst gebruikt oorspronkelijke onderzoeken en technische documentatie. Getallen met een fictieve configuratie zijn didactische berekeningen. Foto's beschrijven

waarnemingen en verwerking; generatieve beelden zijn als illustratie herkenbaar en afzonderlijk gecrediteerd. AXI is een fictieve glazen octaëder met een onderzoeksrol, zonder bewustzijnsclaim. Bij je eigen vergelijking helpen originele

bestanden, modelversie, instellingen en vooraf gekozen criteria om een resultaat terug te vinden. Houd zichtbaar welke informatie werd gemeten en welke door een model is ingevuld.

Bronnen

- 1 LeCun et al. • Gradient-Based Learning (1998) · ieeexplore.ieee.org/document/726791
- 2 Krizhevsky, Sutskever & Hinton • AlexNet (2012) · [cave.cs.toronto.edu/kriz/imagenet_classification_with_deep_convolutional.pdf](https://toronto.edu/kriz/imagenet_classification_with_deep_convolutional.pdf)
- 3 He et al. • Deep Residual Learning (2015) · [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)
- 4 Dosovitskiy et al. • Vision Transformer (2020) · [arXiv:2010.11929](https://arxiv.org/abs/2010.11929)
- 5 Radford et al. • CLIP (2021) · [arXiv:2103.00020](https://arxiv.org/abs/2103.00020)
- 6 MIT • Signals and Systems, Sampling (2011) · web.mit.edu/6.003/F11/www/handouts/lec22.pdf
- 7 Radford et al. • Whisper (2022) · [arXiv:2212.04356](https://arxiv.org/abs/2212.04356)
- 8 Baevski et al. • wav2vec 2.0 (2020) · [arXiv:2006.11477](https://arxiv.org/abs/2006.11477)
- 9 Graves et al. • Connectionist Temporal Classification (2006) · cs.toronto.edu/~graves/icml_2006.pdf
- 10 Rohrbach et al. • Object Hallucination (2018) · [arXiv:1809.02156](https://arxiv.org/abs/1809.02156)
- 11 Ho, Jain & Abbeel • Denoising Diffusion (2020) · [arXiv:2006.11239](https://arxiv.org/abs/2006.11239)
- 12 Nichol & Dhariwal • Improved DDPM (2021) · [arXiv:2102.09672](https://arxiv.org/abs/2102.09672)
- 13 Rombach et al. • Latent Diffusion (2021) · [arXiv:2112.10752](https://arxiv.org/abs/2112.10752)

- 14 Ho & Salimans • Classifier-Free Guidance (2022) · [arXiv:2207.12598](https://arxiv.org/abs/2207.12598)
- 15 Goodfellow et al. • Generative Adversarial Networks (2014) · [arXiv:1406.2661](https://arxiv.org/abs/1406.2661)
- 16 Kingma & Welling • Auto-Encoding Variational Bayes (2013) · [arXiv:1312.6114](https://arxiv.org/abs/1312.6114)
- 17 Ho et al. • Video Diffusion Models (2022) · [arXiv:2204.03458](https://arxiv.org/abs/2204.03458)
- 18 Mildenhall et al. • NeRF (2020) · [arXiv:2003.08934](https://arxiv.org/abs/2003.08934)
- 19 Kerbl et al. • 3D Gaussian Splatting (2023) · [arXiv:2308.04079](https://arxiv.org/abs/2308.04079)
- 20 Geirhos et al. • Texture Bias (2018) · [arXiv:1811.12231](https://arxiv.org/abs/1811.12231)
- 21 Goodfellow, Shlens & Szegedy • Adversarial Examples (2014) · [arXiv:1412.6572](https://arxiv.org/abs/1412.6572)
- 22 Wang et al. • Structural Similarity (2004) · cns.nyu.edu/pub/lcv/wang03-reprint.pdf
- 23 Heusel et al. • Two Time-Scale Update Rule (2017) · [arXiv:1706.08500](https://arxiv.org/abs/1706.08500)
- 24 C2PA • Content Credentials, specificatie 2.2 · spec.c2pa.org/specifications/2.2/specs/C2PA_Specification.html
- 25 Gebru et al. • Datasheets for Datasets (2018) · [arXiv:1803.09010](https://arxiv.org/abs/1803.09010)
- 26 Mitchell et al. • Model Cards (2018) · [arXiv:1810.03993](https://arxiv.org/abs/1810.03993)
- 27 Kirillov et al. • Segment Anything (2023) · [arXiv:2304.02643](https://arxiv.org/abs/2304.02643)

- 28 Vaswani et al. • Attention Is All You Need (2017) · [arXiv:1706.03762](https://arxiv.org/abs/1706.03762)
- 29 Kingma et al. • Variational Diffusion Models (2021) · [arXiv:2107.00630](https://arxiv.org/abs/2107.00630)
- 30 Hamamatsu • Calculating SNR · camera.hamamatsu.com/us/en/learn/technical_information/thechnical_guide/calculating_snr.html
- 31 Kanade e.a. • Virtualized Reality: Concepts and Early Results · cs.cmu.edu/~virtualized-reality/papers/RepVis/RepVis.html
- 32 Geiger, Lenz & Urtasun (2012) • Are we ready for Autonomous Driving? · cvlibs.net/publications/Geiger2012CVPR.pdf
- 33 Ranftl, Bochkovskiy & Koltun (2021) • Vision Transformers for Dense Prediction · [arXiv:2103.13413](https://arxiv.org/abs/2103.13413)
- 34 Anthropic • Claude Opus 5.5, officiële modeldocumentatie · platform.claude.com/docs/en/models/opus-5-5/overview
- 35 Anthropic • Opus-benchmarks en evaluatievoetnoten · anthropic.com/claude/opus

Landscape of Star Birth · NASA, ESA, CSA, STScI / Webb ERO Production Team · Publiek domein · Infraroodopname in toegewezen zichtbare kleuren | p. 4 · A waterfall in the forest – geograph.org.uk – 6574048 · David Medcalf · CC BY-SA 2.0

Illustraties

p. 1 • Vector 03 cover new · Originele AI-illustratie voor VECTOR | p. 2, 17 · AXI en de auto zonder achterkant · Originele AI-illustratie voor VECTOR | p. 2, 3 · VECTOR 03 — zesdelige strip · Originele AI-illustratie voor VECTOR | p. 2, 22, 23 · VECTOR 03 — conceptillustratie voor dubbele pagina · Originele AI-illustratie voor VECTOR | p. 2, 12 · Vector 03 stereo depth v2 · Originele AI-illustratie voor VECTOR | p. 6, 7 · VECTOR 03 — conceptillustratie voor dubbele pagina · Originele AI-illustratie voor VECTOR | p. 7, 23 · AXI, de waarnemer · Originele AI-illustratie voor VECTOR | p. 19 · VECTOR 03 — zesdelige strip · Originele AI-illustratie voor VECTOR | p. 50 · VECTOR 04 — omslagillustratie · Originele AI-illustratie voor VECTOR

Foto's zijn verkleind en kunnen voor de opmaak zijn bijgesneden; ze houden hun aangegeven licentie. Teksten, diagrammen, puzzels en vormgeving: VECTOR.

Foto's

Via Wikimedia Commons, tenzij anders vermeld.

p. 2, 43 • Daylight through forest 001 · BVBurton · CC BY-SA 4.0 | p. 2, 38 • NASA's Webb Reveals Cosmic Cliffs, Glittering



IN DE VOLGENDE EDITIE • Vanaf 8 oktober 2026

Betrouwbare AI in de praktijk

- Maak van een score een controleerbare test
- Onderzoek onzekerheid, fouten en ontbrekende informatie
- Bewaak de grens tussen advies en uitvoering