KU LEUVEN

FACULTEIT ECONOMIE EN
BEDRIJFSWETENSCHAPPEN
CAMPUS BRUSSEL

Portrait Matting in de Google Pixel 6

Artikel: ‘The Pixel 6's selfie portrait mode can see individual strands of hair, and here's how it works’ (Hager, 2022)

Introductie

In deze essay tracht ik een kader te vormen voor de nieuwe image matting techniek die Google implementeert in hun Google Pixel 6, namelijk de Portrait-Matting techniek. Dit doe ik aan de hand van het artikel van Ryne Hager (2022) van 'androidpolice.com', waarin hij de techniek en de dataverzameling kort samenvat. Om de relevantie van deze techniek goed te begrijpen en kaderen binnen de reeks Pixel smartphones van Google zelf, heb ik de evolutie en context van de portretmodus functie kort aangehaald a.d.h.v. de Google AI blog. Echter is de relevantie hiervan voor de AI-techniek in kwestie iets minder belangrijk, wat mij er heeft toe aangezet deze sectie naar appendix A te verplaatsen. Het algemene onderwerp dat ik aanhaal is 'Portrait-Matting', een zelfontworpen techniek die Google implementeert om de kwaliteit van de portretmodus in de Pixel 6 smartphone te optimaliseren. De portretmodus is namelijk een camera-functie waarbij de persoon (of een object) wordt onderscheiden van de achtergrond, waarbij de achtergrond een wazige schijn verkrijgt. De manier waarop de scheiding tussen voor -en achtergrond bepaald wordt is iets dat in een smartphone niet natuurlijk gebeurt, waardoor bedrijven zich via deze camera-feature zoveel mogelijk proberen te onderscheiden van hun concurrentie door innovatieve manieren te vinden om een professionele portretfoto te genereren.

Pixel 6

Wat de recente Google Pixel 6 zo speciaal maakt is dat het de eerste Pixel smartphone is met een mobiele system-on-a-chip[1], namelijk de Google Tensor Chip. Hierbij zijn er verschillende verwerkings componenten die de specifieke Machine Learning modellen van Google een optimale werking bieden (Gupta & White, 2021). Een voorbeeld hiervan is NAS (Neural Architecture Search, een subsectie van AutoML), waarmee het mogelijk is om de meest efficiënte en effectieve Machine Learning technieken te vinden/ontwikkelen binnen het bereik van enkele parameters (zoals een vastgestelde parameter op vlak van latentie[2] en stroomvereisten). Het hebben van dergelijke componenten brengt verschillende mogelijkheden voor het toestel (en vooral de camera) met zich mee, de zoekruimte dat beschikbaar wordt door deze NAS maakt het bijvoorbeeld mogelijk een functie te creëren op de smartphone om wazige pixels over een gezicht scherp te maken (Gupta & White, 2021).

Voor de portret modus is er ook een belangrijke nieuwigheid, namelijk de implementatie van '**Image Matting**' (Escolano & Ehman, 2022). Deze techniek verwijst naar het proces waarbij men verschillende lagen aan pixels in een afbeelding onderscheidt, waarbij meestal gewerkt wordt met slechts twee lagen, namelijk een voorgrond en een achtergrond. Deze techniek verschilt met de '**segmentatie**' techniek (Appendix A) in dat er bij segmentatie op een binaire manier tewerk wordt gegaan, oftewel behoort een pixel tot de voorgrond of tot de achtergrond (het vinden van grenzen is hierbij echter niet altijd accuraat, o.a. zoals bevonden door Chen en Pavlidis, 1980), terwijl bij **alpha mattes** (een onderdeel van het Image Matting, waarmee men de verschillende lagen 'afbeeldingen' die in een afbeelding worden teruggevonden

[1] 'een geïntegreerde schakeling (IC) die alle componenten van een computer of elektronisch systeem samenvoegt in de behuizing van een enkele chip (Wikipedia-bijdragers, 2021).

[2] Latency is hetzelfde als 'lag', waarmee een vertraging in de dataoverdracht over het datacommunicatienetwerk wordt bedoeld (Wikipedia-bijdragers, 2022).

bedoeld) er een veel preciezere meting gebeurt om de randen van een object te bepalen. Het doel bij Image matting is dan ook om de transparantie te meten van de voorgrond (het in-focus vlak van de portretmodus) aan de hand van een 'trimap[3]', waarbij de onzekere randen gemeten moeten worden (Cho et al., 2019). In de Pixel 6 smartphone is er gebruik gemaakt van een nieuwe techniek die de alpha matte accurater vindt en een hogere resolutie bezit.

AI-techniek

De techniek dat hierbij gebruikt wordt is door Google bestempeld als '**Portrait Matting**', waarbij men aan de hand van een fully convolutional neural network (FCN) geleidelijk een uiterst gedetailleerde Alpha Matte kan schatten (Escolano & Ehman, 2022).

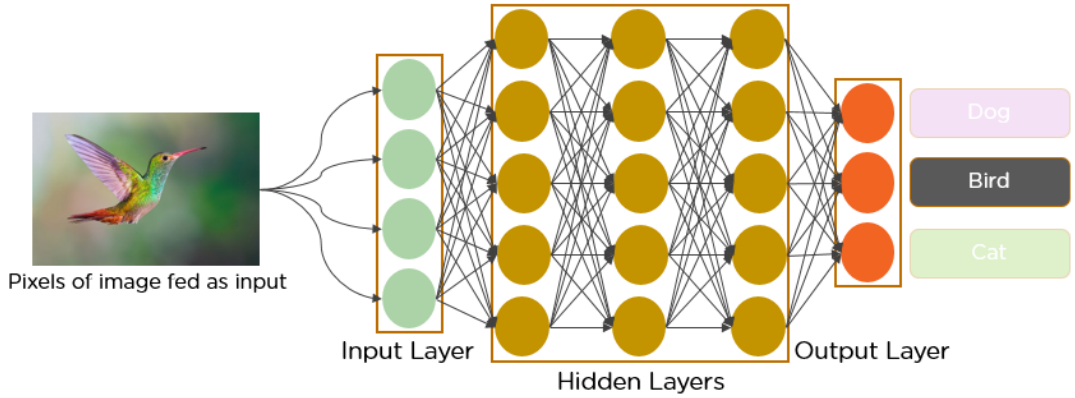
Een Neuraal Netwerk (NN) wordt best beschreven als een netwerk van verschillende processoren (neuronen) die met elkaar verbonden zijn. Input-neuronen worden initieel geactiveerd door sensoren die een omgeving innemen, waarna andere neuronen geactiveerd worden als opvolging van initiële neuronen via gewichten die toegekend zijn (Schmidhuber, 2015). Zoals we al eerder ondervonden in het MOOC[4] (elements of AI) bestaat een convolutioneel neuraal netwerk (CNN) eruit dat er een convolutionele laag wordt opgenomen in een diep neuraal netwerk. Zo'n CNN is een deep learning algoritme waarbij er als input een afbeelding is en men via het toekennen van gewichten aan eender welk aspect uit de afbeelding zaken kan classificeren/herkennen en van elkaar onderscheiden (Saha, 2021). Een CNN heeft (meestal) namelijk als doel bepaalde hersenactiviteiten na te bootsen en doet dit door de neuronen op een specifieke manier te organiseren (Wat is een convolutioneel neuraal netwerk?, z.d.; Data Science Team, 2020a). In het MOOC werd de nood en de functie van een CNN aangehaald a.d.h.v. enkele voorbeelden. Denk hierbij bijvoorbeeld aan de classificatie van een smiley, waarbij men nood had aan een CNN die de classificatie mogelijk maakte via de verschillende lagen. Door een toevoeging van meerdere lagen en het gebruik van backpropagation[5] kan men in combinatie met een CNN (wegens de grote trainingset) namelijk op een veel accuratere manier objecten in afbeeldingen herkennen. Een verder voorbeeld uit het MOOC is dat van het verkeersbord: om een verkeersbord aan de hand van een neuraal netwerk te herkennen in een afbeelding (op basis van pixels) heeft men een enorme hoeveelheid trainingsgegevens nodig waarbij het verkeersbord op exact dezelfde positie (zelfde pixels) voorkomt. Indien men in tegenstelling tot een gewoon neuraal netwerk een CNN gebruikt, zou men het verkeersbord op eender welke positie in de afbeelding kunnen

[3] Een ruwe segmentatie van een afbeelding waarbij onderscheid wordt gemaakt tussen drie regio's, de zekere voorgrond, de zekere achtergrond en de onzekere regio waarbij er een overgang is tussen voor -en achtergrond (Semkin, 2021).

[4] Massive Online Open Courses (elements of AI).

[5] Een algoritme dat gebruikt wordt voor de ontwikkeling van 'Feedforward Neural Networks' (Wikipedia contributors, 2022a). In een kunstmatig neuraal netwerk wordt een signaal ontvangen (input layer in feedforward NN), verwerkt en getraind, waarna het doorgaat naar verborgen neuronen (hidden layer in feedforward NN). De verbinding van het ene neuron met de andere wordt bepaald door de hoeveelheid gewicht dat eraan wordt toegekend, waarbij de toename/afname van het signaal van een neuron afhankelijk is van het gewicht dat eraan is toegekend. Dat signaal gaat dan door naar uitgangs-neuronen (output layer in feedforward NN) (Wikipedia contributors, 2022b; Data Science Team, 2020b) Hieraan wordt verder toegevoegd: 'Wanneer kunstmatige neurale netwerken worden gevormd, worden de waarden van de gewichten willekeurig toegewezen. Wanneer de waarde afwijkt van het verwachte feedforward netwerk, is er een fout. Het algoritme wordt zo ingesteld dat het model de parameters verandert telkens als de output niet de verwachte is. Als de parameter verandert, verandert de fout ook totdat het neurale netwerk de gewenste output vindt door de gradiënt daling te berekenen.' (Data Science Team, 2020b).

herkennen zonder dat er voldaan moet worden aan de eisen van het klassieke neurale netwerk (bezitten over trainingsgegevens met identieke pixel-specificaties (grootte, plaats,…))



(CNN for Deep Learning | Convolutional Neural Networks, 2021)

Bij de Pixel 6 gaat het echter om een fully convolutional neural network (FCN). Het verschil tussen een FCN en een CNN zit zich in de structuur en opmaak van de lagen, bij een standaard CNN bestaat de laatste laag voor de output uit een 'dense-layer' (of 'fully connected layers), terwijl bij FCN de 'dense-layers' telkens vervangen zijn door convolutionele lagen (Fu et al., 2017). Een normale CNN heeft het nadeel dat modellen voor beeldclassificatie en objectdetectie regelmatig getraind zijn op basis van vaste beeldformaten. Indien afbeeldingen uit de trainingset niet dezelfde afmetingen hebben worden deze aangepast, wat o.a. tot gevolg kan hebben dat het herschalen van een afbeelding belangrijke kenmerken aantast (Rawlani, 2021). In tegenstelling tot een CNN kan men bij een FCN compacte (dense) outputs leveren van inputs die eender welke grote/beeldformaten bevatten (Shelhamer et al., 2017). Deze feature kan de FCN danken aan de knelpunt-lagen die de twee essentiële componenten (encoder en decoder) van het netwerk verbinden (Henry et al., 2018).

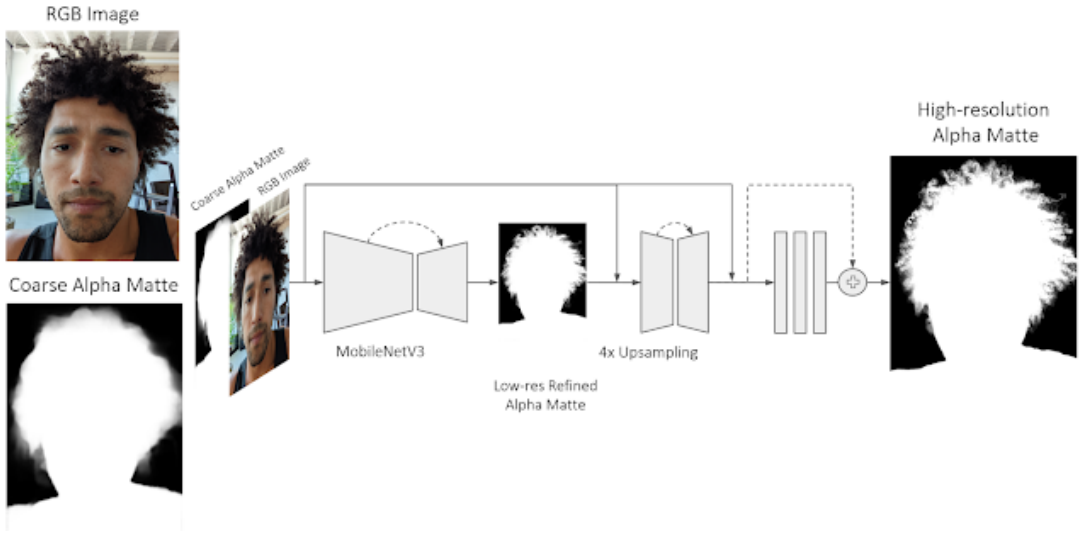
De FCN gebruikt een MobileNetV3 architectuur en bestaat uit een reeks 'encoder-decoder blocks'. De encoder-component bestaat eruit dat de input met variabele vorm omgevormd wordt tot één vaste vorm (Zhang et al., 2021), de beelden worden namelijk geanalyseerd en omgevormd tot een cluster van voorspellingen, data uit afbeeldingen worden dan geleidelijk gedownsampled tot ze proportioneel betekenisvoller worden (Henry et al., 2018). De decoder-component gebruikt vervolgens upsampling operaties om de ruimtelijke eigenschappen te herstellen van de voorspellingen tot deze een niveau bereiken waar ze dezelfde grootte bevatten als de input afbeelding (Henry et al., 2018).

De eerste versie van de eerder vernoemde MobileNetV3 was de MobileNets, een computervisie model[6] opgesteld om op een smartphone te functioneren. Het gaat om kleine modellen binnenin een smartphone die voldoet aan enkele essentiële parameters (wegens de beperkte hardware capaciteiten van een smartphone) waarop men verschillende features kan bouwen zoals o.a. segmentation en classification (Howard & Zhu, 2017). Om de parameters waaraan dergelijke architectuur moet voldoen zo nuttig mogelijk te handhaven, maakt men bij

[6] Het vertalen van visuele data op basis van functies en contextuele informatie die men geïdentificeerd heeft tijdens het trainen van het model (Deep Learning for Computer Vision, z.d.).

MobileNetV3 in de PIxel 6 gebruik van AutoML (zie neural architecture search aan het begin van deze paragraaf), waardoor het mogelijk wordt de beste architectuur te vinden binnen de benodigde parameters. De implementatie van deze MobileNetV3 zorgt voor een gelijkaardige accuraatheid als de vorige versie (MobileNetV2) maar behaalt de resultaten dubbel zo snel (Howard & Gupta, 2019).

In de Pixel 6 smartphone wordt er als eerste stap een selfie genomen met de portretmodus, waarbij er een initieel ruw masker gemaakt wordt van de gezichts-uitlijning (coarse alpha matte[7]), en zowel de volledige selfie (kleurenfoto) als de ruwe alpha matte in het model worden gezet (Hager, 2022). Portrait matting past dan een ondiepe decoder (lage hoeveelheid lagen) toe op deze input waardoor men een masker krijgt dat zeer gedefinieerd is, maar weinig resolutie[8] heeft. Om dit masker vervolgens een hogere resolutie te geven wordt er net zoals eerder aangegeven gebruik gemaakt van een ondiepe encoder-decoder (Escolano & Ehman, 2022).

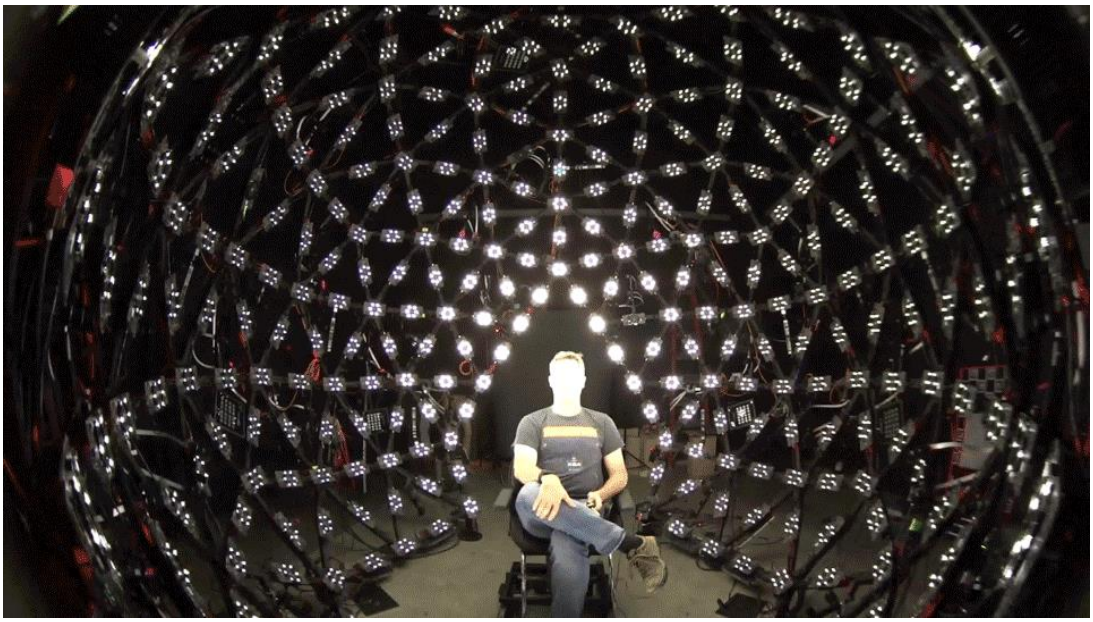


Het encoder-decoder proces bij Portrait Matting in de Pixel 6 (Escolano & Ehman, 2022)

[7] Opgesteld door een lage resolutie personen segmenter zoals bij andere smartphones.
[8] In de fotografie verwijst men hiermee naar de beeldscherpte (resolutie, z.d.).

Type data

Modellen zoals Portrait matting zijn echter enkel effectief als men deze kan opstellen met een enorme hoeveelheid diverse en accurate trainingsgegevens, waarbij personen in eender welk licht afgebeeld staan en van eender welke kant gefotografeerd worden (Hager, 2022). Om de verschillende uitdagingen die hierbij komen kijken te weerstaan maakt Google gebruik van 'Ground Truth Data' (Escolano & Ehman, 2022), waarmee men simpelweg verwijst naar data die representatief is voor de werkelijkheid. Om dit type data te verkrijgen maakt men gebruik van een 'volumetric capture system' , waarbij een persoon in een bolvormige structuur wordt geplaatst die volhangt met camera's en belichtingssystemen (64 camera's en 331 individueel programmeerbare LED lichten). De persoon wordt vervolgens gefotografeerd met ieder individueel licht apart (OLAT, One light at a time) waardoor de persoon's uiterlijk verkregen wordt vanuit ieder lichtinval punt. Als men de afbeeldingen dan allemaal samenvoegd krijgt men het uiterlijk dat verkregen kan worden met eender welk beeld gebaseerde verlichting[9] (Tsai & Pandey, 2020).



Voorbeeld van een Volumetric Capture System (Hager, 2022)

Eén van de manieren waarop de personen (trainingset) gefotografeerd wordt in deze structuur is met alle lichten uit, maar met een verlichte achtergrond, wat het mogelijk maakt alle individuele haren tegen dit oppervlak te identificeren. Aan de hand van Light stage[10] creëert/berekent men voor deze personen vervolgens een Alpha Matte, die men via een deep learning techniek mede berekent voor iedere foto (met verschillende lichtinval en standpunten) uit de structuur. Hierna kan men de persoon a.d.h.v. deze verschillende alpha mattes positioneren in eender welke achtergrond, waarbij de belichting op een realistische manier weergegeven wordt (Escolano & Ehman, 2022).

[9] (Image based lighting)

[10] Een computationeel belichtingssysteem in de Volumetric Captur System (Tsai & Pandey, 2020).

Resultaat van het creëren van fotosets voor de trainingset (Escolano & Ehman, 2022)

Met deze training set die men zelf gecreëerd heeft, heeft men een basis aan zeer betrouwbare gegevens die men vervolgens combineert met verschillende echte foto's in natuurlijke achtergronden om tot de volledige trainingset van het Portrait matting model te komen. Deze toevoeging van foto's die mensen nemen in natuurlijke achtergronden worden via een model met hoge accuraatheid omgezet in alpha mattes. Wegens het trainen van dit Portrait Matting model o.b.v. diverse data is het dus geschikt om accurate alpha mattes te schatten in een brede reeks aan omstandigheden (type mensen, achtergronden, haarstijlen,...) (Escolano & Ehman, 2022; Hager, 2022).

Doelgroep van techniek en Maatschappelijke impact

Het gebruik van deze techniek in de portretmodus van de Pixel 6 brengt voordelen mee voor iedereen, gaande van de individuele gebruiker tot de sector als geheel. De functie maakt het heel gebruiksvriendelijk naar zeer brede groep mensen toe en maakt het mogelijk voor mensen met speciale haar -en kledingstijlen om goede gebruikservaring te beleven met de portretmodus. Zoals al werd aangegeven is het meestal niet eenvoudig om de overgang in foto's tussen voor -en achtergrond te meten/bepalen, maar met deze nieuwe 'Portrait Matting' techniek is er een definitieve stap gezet in de richting van excellentie, wat niet uitsluit dat er nog ruimte is voor verbetering. Ryne Hager (2022) wijst in zijn artikel op één van de minpunten waar nog verbetering nodig is, namelijk het verloop van de wazigheid, die nog steeds niet even scherp is als dat van een digitale spiegelreflexcamera (DSLR, denk bijvoorbeeld aan een Canon EOS 2000D).

Ook de sector als geheel haalt enorme voordelen van de technieken die worden ontwikkeld bij Google. Google maakt er namelijk een gewoonte van om geïmplementeerde technieken open source te maken, wat wil zeggen dat eender welke persoon naar de code kan kijken, deze kan veranderen, enz. Dit open-source aspect gebeurt via Tensorflow, een platform dat specifiek ontwikkeld is voor open source machine learning technieken. Escolano en Ehman (2022) wijzen er verder ook op dat veel van het hedendaagse deep learning werk in image matting bestaat uit technieken die onmogelijk toepasbaar zijn op enorme datasets, hetgeen waar Google gebruik van maakt. Om de problemen die deze hedendaagse technieken met zich meebrengen (en de problemen die verder uit deze technieken voortvloeien) te vermijden is Google hun eigen image matting techniek gestart (portrait matting), wat hun leiderschapspositie in de sector nogmaals benadrukt.

Ethische en wettelijke aspecten

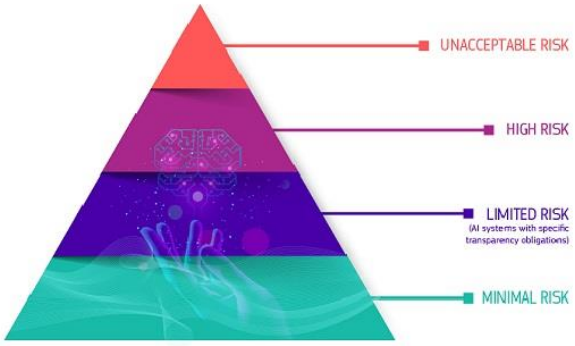
Ook op ethisch vlak streeft Google naar excellentie en verantwoordelijkheid. Dit is opmerkelijk zichtbaar bij de benadrukking van de 'AI-Principles' (Google AI, z.d.-a), die bij iedere Pixel Smartphone in het achterhoofd worden gehouden voor de implementatie van nieuwe Artificiële Intelligentie features. Bij de principes die Google centraal zet vindt men onder andere het belang aan maatschappelijke voordelen, het vermijden van oneerlijke vooroordelen (bijvoorbeeld op basis van geslacht, ras,...) en de handhaving van principes omtrent privacy.

Net zoals bij college van Professor Smuha licht AI4Belgium (2019) de 7 vereisten[11] van betrouwbare AI toe die voorgelegd werden door de Europese Commissie. Aan elk van deze vereisten ben ik van mening dat Google voldoet en dit wil ik aantonen aan de hand van de controle -en toezicht voorwaarde, de non-discriminatie voorwaarde en de transparantie voorwaarde (wat niet wegneemt dat de andere voorwaarden niet evenwaardig behandeld worden bij Google, zoals zichtbaar in hun AI-Principles). Voor het beantwoorden aan de controle -en toezicht voorwaarde is er bij Google een opgesteld team dat verantwoordelijk is voor ethische reviews, waarbij Google heel transparant is en een vast proces doorloopt om de uitdagingen aan te gaan (Google AI, z.d.-b). De transparantie voorwaarde wordt dan weer beantwoord door de gecreëerde modellen open source beschikbaar te stellen op TensorFlow

[11] 1) menselijke controle en toezicht, (2) technische robuustheid en veiligheid, (3) privacy en datagovernance, (4) transparantie, (5) diversiteit, non-discriminatie en eerlijkheid, (6) maatschappelijk en milieuwelzijn en (7) verantwoording (AI4Belgium, 2019).

(zoals eerder besproken). Hiernaast wordt er om bedachtzaam om te gaan met zaken als discriminatie en vooroordelen in het specifieke geval van Portrait Matting in de Pixel 6 voor gezorgd dat het model getraind is op basis van een heel diverse dataset (huidskleuren, haarstijlen, kledingstijlen,...), waarmee men de vernoemde bias uit het college[12] zo goed mogelijk vermijdt.

Wegens de natuur en context van de Portrait matting techniek ben ik er tevens van overtuigd dat Google in het risico-gebaseerde wettelijke kader (de voorgestelde 'AI-act') behoort tot de sectie 'minimaal of geen risico' of 'limited risk'. Er kan met zekerheid gezegd worden dat Portrait Matting niet past binnen de hogere of onacceptabele risico bepalingen, aangezien men in dergelijke categorieën zaken terugvindt die o.a. de gezondheid, kansen, en algemene menselijke rechten beïnvloeden. De techniek die beschreven werd in het artikel van Hager (2022) is dit zeker niet, en biedt zelfs een veel inclusievere mogelijkheid tot succesvolle portretfoto's. Ook onder de sectie 'gelimiteerd risico' kan deze techniek niet geplaatst worden, aangezien men in deze sectie AI-toepassingen terugvindt die '(i) interactie hebben met mensen, (ii) worden gebruikt om emoties te detecteren of associatie met (sociale) categorieën te bepalen op basis van biometrische gegevens, of (iii) inhoud genereren of manipuleren ("deep fakes")' (European Commission, 2021). Portrait Matting valt niet te plaatsen onder deze drie aangehaalde punten en plaats ik bijgevolg in de laatste sectie van het wettelijk kader, namelijk 'minimaal risico', waar geen restricties voor worden voorgesteld.



Een risico-gebaseerde wettelijke benadering van AI-technieken (European Commission, 2022)

Conclusie

Door het gebruik van de zelf-ontworpen Portrait Matting (a.d.h.v. een fully convolutional neural network) is Google erin geslaagd een (bijna) perfecte portret modus na te bootsen in de Pixel 6 smartphone. Een techniek die voor de meeste partijen die er mee in aanraken komen een volkomen voordeel biedt en zowel op ethisch als juridisch vlak een opmerkelijke mijlpaal bereikt.

[12] Namelijk dat het AI-model voornamelijk getraind is o.b.v. training sets van blanke gezichten, waardoor de classificatie/herkenning van mensen met een donkere huidskleur niet altijd correct gebeurt.

Bronnen

AI4Belgium. (2019, 18 april). Europese Commissie presenteert 'Ethische Guidelines voor Artificiële Intelligentie'. Geraadpleegd op 27 april 2022, van https://www.ai4belgium.be/nl/europese-commissie-presenteert-ethische-guidelines-voor-artificiele-intelligentie/#:%7E:text=Na%20de%20analyse%20van%20ethische

Chen, & Pavlidis, T. (1980). Image segmentation as an estimation problem. Computer Graphics and Image Processing, 12(2), 153–172. https://doi.org/10.1016/0146-664X(80)90009-X

Cho, Tai, Y.-W., & Kweon, I. S. (2019). Deep Convolutional Neural Network for Natural Image Matting Using Initial Alpha Mattes. IEEE Transactions on Image Processing, 28(3), 1054–1067. https://doi.org/10.1109/TIP.2018.2872925

CNN for Deep Learning | Convolutional Neural Networks. (2021, 23 juli). Analytics Vidhya. Geraadpleegd op 25 april 2022, van https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

Data Science Team. (2020a, 18 december). Convolutionele neurale netwerken – de ELI5 manier. DATA SCIENCE. Geraadpleegd op 24 april 2022, van https://datascience.eu/nl/computer-vision/convolutionele-neurale-netwerken-de-eli5-manier/

Data Science Team. (2020b, 18 december). De eenvoudige uitleg van het concept van backpropagatie. DATA SCIENCE. Geraadpleegd op 24 april 2022, van https://datascience.eu/nl/kunstmatige-intelligentie/hoe-het-algoritme-voor-rugpropagatie-werkt/

Deep Learning for Computer Vision. (z.d.). Run.ai. Geraadpleegd op 25 april 2022, van https://www.run.ai/guides/deep-learning-for-computer-vision#:%7E:text=Computer%20vision%20models%20are%20designed,predictive%20or%20decision%20making%20tasks.

Escolano, S. O., & Ehman, J. (2022, 24 januari). Accurate Alpha Matting for Portrait Mode Selfies on Pixel 6. Google AI Blog. Geraadpleegd op 17 april 2022, van https://ai.googleblog.com/2022/01/accurate-alpha-matting-for-portrait.html

European Commission. (2021, 21 april). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. EUR.Lex. Geraadpleegd op 29 april 2022, van https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

European Commission. (2022, 30 maart). Regulatory framework proposal on artificial intelligence. Shaping Europe's Digital Future. Geraadpleegd op 29 april 2022, van https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017). Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. Remote Sensing, 9(5), 498. https://doi.org/10.3390/rs9050498

Garg, R., & Wadhwa, N. (2018, 29 november). Learning to Predict Depth on the Pixel 3 Phones. Google AI Blog. Geraadpleegd op 13 april 2022, van https://ai.googleblog.com/2018/11/learning-to-predict-depth-on-pixel-3.html

Google AI. (z.d.-a). Our Principles. Geraadpleegd op 27 april 2022, van https://ai.google/principles/

Google AI. (z.d.-b). Review Process. Geraadpleegd op 27 april 2022, van https://ai.google/responsibilities/review-process/

Gupta, S., & White, M. (2021, 8 november). Improved On-Device ML on Pixel 6, with Neural Architecture Search. Google AI Blog. Geraadpleegd op 15 april 2022, van https://ai.googleblog.com/2021/11/improved-on-device-ml-on-pixel-6-with.html

Hager, R. (2022, 1 februari). The Pixel 6's selfie portrait mode can see individual strands of hair, and here's how it works. Android Police. Geraadpleegd op 14 april 2022, van https://www.androidpolice.com/pixel-6-selfie-portrait-mode/

Henry, C., Azimi, S. M., & Merkle, N. (2018). Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. IEEE Geoscience and Remote Sensing Letters, 15(12), 1867–1871. https://doi.org/10.1109/LGRS.2018.2864342

Hirokawa, N., & Windhorst, U. (2008). Aperture Problem. Encyclopedia of Neuroscience, 159. https://doi.org/10.1007/978-3-540-29678-2_310

Howard, A. G., & Zhu, M. (2017, 14 juni). MobileNets: Open-Source Models for Efficient On-Device Vision. Google AI Blog. Geraadpleegd op 26 april 2022, van https://ai.googleblog.com/2017/06/mobilenets-open-source-models-for.html

Howard, A., & Gupta, S. (2019, 13 november). Introducing the Next Generation of On-Device Vision Models: MobileNetV3 and MobileNetEdgeTPU. Google AI Blog. Geraadpleegd op 26 april 2022, van https://ai.googleblog.com/2019/11/introducing-next-generation-on-device.html

Levoy, M., & Pritch, Y. (2017, 17 oktober). Portrait mode on the Pixel 2 and Pixel 2 XL smartphones. Google AI Blog. Geraadpleegd op 11 april 2022, van https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html

Matcha, A. C. N. (2021, 20 mei). A 2021 guide to Semantic Segmentation. AI & Machine Learning Blog. Geraadpleegd op 11 april 2022, van https://nanonets.com/blog/semantic-image-segmentation-2020/

Poot, R. (2022, 1 maart). Wat is DOF? Natuurfotografie. Geraadpleegd op 11 april 2022, van https://www.natuurfotografie.nl/wat-is-dof/

Rawlani, H. (2021, 13 december). Understanding and implementing a fully convolutional network (FCN). Towards Data Science. Geraadpleegd op 25 april 2022, van https://towardsdatascience.com/implementing-a-fully-convolutional-network-fcn-in-tensorflow-2-3c46fb61de3b

Resolutie. (z.d.). Van Dale. Geraadpleegd op 26 april 2022, van https://www.vandale.nl/gratis-woordenboek/nederlands/betekenis/resolutie#.YmhChNpBw2w

Saha, S. (2021, 7 december). A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Medium. Geraadpleegd op 25 april 2022, van https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Schmidhuber. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Semkin, P. (2021, 3 mei). Image Matting with state-of-the-art Method “F, B, Alpha Matting” | LearnOpenCV. LearnOpenCV – OpenCV, PyTorch, Keras, Tensorflow Examples and Tutorials. Geraadpleegd op 24 april 2022, van https://learnopencv.com/image-matting-with-state-of-the-art-method-f-b-alpha-matting/

Shelhamer, Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

Tsai, Y., & Pandey, R. (2020, 11 december). Portrait Light: Enhancing Portrait Lighting with Machine Learning. Google AI Blog. Geraadpleegd op 27 april 2022, van https://ai.googleblog.com/2020/12/portrait-light-enhancing-portrait.html

Vogelaar, E. (2018, 11 juli). Wat is hdr en wat heb je eraan op je smartphone? Android Planet. Geraadpleegd op 13 april 2022, van https://www.androidplanet.nl/nieuws/wat-is-hdr-uitleg/

Wadhwa, N., & Zhang, Y. (2019, 16 december). Improvements to Portrait Mode on the Google Pixel 4 and Pixel 4 XL. Google AI Blog. Geraadpleegd op 14 april 2022, van https://ai.googleblog.com/2019/12/improvements-to-portrait-mode-on-google.html

Wat is een convolutioneel neuraal netwerk? (z.d.). Netinbag. Geraadpleegd op 24 april 2022, van https://www.netinbag.com/nl/internet/what-is-a-convolutional-neural-network.html

Wikipedia contributors. (2022a, 11 maart). Backpropagation. Wikipedia. Geraadpleegd op 24 april 2022, van https://en.wikipedia.org/wiki/Backpropagation

Wikipedia contributors. (2022b, februari 10). Feedforward neural network. Wikipedia. Geraadpleegd op 24 april 2022, van https://en.wikipedia.org/wiki/Feedforward_neural_network

Wikipedia-bijdragers. (2021, 19 mei). System-on-a-chip. Wikipedia. Geraadpleegd op 5 april 2022, van https://nl.wikipedia.org/wiki/System-on-a-chip

Wikipedia-bijdragers. (2022, 25 februari). Latentie. Wikipedia. Geraadpleegd op 17 april 2022, van https://nl.wikipedia.org/wiki/Latentie#:%7E:text=Latentie%20

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning.

Appendix

Appendix A: Evolutie en context van de portretmodus (in de Pixel smartphone reeks)

Levoy en Pritch (2017) geven in de Google AI-blog weer hoe een oorspronkelijke camera met een grote lens (zoals bijvoorbeeld de Spiegelreflexcamera SLR) een portrait-effect creëert. Dergelijke spiegelreflexcamera's hadden een grote lens en werkten met een nauwe/ondiepe diepteschерpte[13], alles wat ofwel voor of achter het vlak dat 'in-scherpte' ligt krijgt een wazige glans. Het 'ondiepe' aspect van deze SLR-camera's is wat hun foto's een artistieke uitstraling geven en de afstand tussen het object dat scherp in beeld wordt gebracht en de achtergrond/voorgrond is hierbij het bepalende voor hoe wazig dat gedeelte van de foto wordt. De camera's die gebruikt worden in GSM's en smartphones werken echter niet op deze manier, waardoor men op een synthetische manier zal moeten bepalen welke delen van de foto op welke afstand liggen van het object. Wanneer men erachter is gekomen wat deze afstanden zijn (door technieken toe te passen zoals het plaatsen van 2 camera's die het effect van ogen nabootsen via een stereo algoritme (dual-camera phones) en **semantische segmentatie**[14]), kan men zowel de wazigheid nabootsen (gemiddelde nemen van de kleur van pixel en zijn buren) als vorm van deze wazigheid (bokeh) die de lens van een SLR-camera creëert.

Aangezien de Pixel 2 portrait modus aanbiedt op zowel de voor als achter camera zijn er verschillende technieken die van toepassing zijn, waarbij zowel semantische segmentatie (zowel voor als achtercamera) en stereo (achtercamera) actief gebruikt worden. Het verhaal bij het gebruik van de achtercamera is echter iets wat complexer aangezien hier een mix gebruikt wordt van de stereo en de semantische segmentatie technieken terwijl er slechts 1 camera aanwezig is (herinner: voor stereo heb je 2 camera's nodig om diepte te meten). Om een lang verhaal kort te beschrijven wordt een stap voor stap te werk gegaan: eerst wordt de afbeelding omgezet in een HDR+ foto (manier om foto's levendiger en kleurrijker te maken (Vogelaar, 2018)), hierna wordt er via een specifieke Machine Learning techniek (convolutional neural network CNN, de betekenis laten we voor de volgende paragraaf) bepaald welk deel van de foto / welke pixels behoren tot de achtergrond en welk deel tot de voorgrond. Naar aanleiding van de mankementen hiervan (bijvoorbeeld het gebrek aan meting van diepte, waarvan men gebruik maakt om de 'hoeveelheid' wazigheid te bepalen), zal men

[13]Scherptediepte (Depth of Field) is de afstand tussen het dichtbijzijnste en het verste punt dat scherp wordt afgebeeld in een foto, bepalend voor de grote van zo'n scherptediepte zijn aspecten als de kwaliteit van de lens, de afstand tot het object, type (camera) sensor en grootte van het diafragma (Poot, 2022).

[14]Een techniek waarbij de pixels van een afbeelding wordt opgesplitst om meerdere afbeeldingen met 'pixel-groepen' te maken. Men classificeert elke pixel dus tot een label (bv. boom, kat,... die in de afbeelding aanwezig zijn) en op deze manier kan men voorgrond en achtergrond onderscheiden (Matcha, 2021). Merk hierbij op dat de hoeveelheid wazigheid niet kan worden gemeten wegens het gebrek aan meting tussen object en achtergrond (Levoy & Pritch, 2017)

op basis van het resultaat van CNN een technologie toepassen dat deze diepte kan meten (deze compenseert voor het gebrek aan meerdere camera's) genaamd 'dual-pixel autofocus' (PDAF). Bij deze PDAF technologie wordt het idee van stereo met 2 camera gereproduceerd door de lens van de achtercamera in twee te splitsen, aangezien de twee helften ongeveer een millimeter van elkaar verwijderd zijn kan men het stereo algoritme toepassen en toch tot een diepte meting komen via een fenomeen genaamd '**parallax**' (wat er gebeurt wanneer je met 2 ogen ergens naar kijkt, je ziet het object van 2 verschillende posities waardoor de verschuiving van het object vanuit de 2 perspectieven een waarneming vormt van de diepte). Ten laatste worden deze stappen allemaal aan elkaar gevoegd afhankelijk van de noden (gebruik voor-camera, achter-camera,…) (Levoy & Pritch, 2017).

Voor de Pixel 3 is Google verdergegaan op de features van de Pixel 2, Garg en Wadhwa (2018) halen aan hoe men de kwaliteit van de diepte-metingen met de dual-pixel autofocus verbeterd wordt in de Pixel 3 met Machine Learning technieken. Een gebrek van de PDAF-techniek is het befaamde 'aperture problem', waarbij de beweging van een eendimensionale structuur niet ondubbelzinnig kan worden bepaald wanneer men ernaar kijkt vanuit een kleine opening (zoals de cameralens) (Hirokawa & Windhorst, 2008). Dit gebrek is wat voor de Pixel 2 regelmatig de oorzaak was van foutmeldingen en slechte dieptemetingen. Om de features van de PDAF-techniek te verbeteren is men ook beroep gaan doen op andere parameters die aanwijzing geven over de diepte van een object. Een van deze parameters is bijvoorbeeld een '**semantische aanwijzing**' (semantic cue). Bij zo'n semantische aanwijzing baseert men zich op de grootte van het object in een afbeelding, wanneer we een afbeelding hebben van iemand die enkele meters in de voorgrond staat van een andere persoon, weten we dat hier een diepteverschil aanwezig is aangezien o.a. het hoofd van deze tweede persoon minder pixels inneemt dan de persoon die in het 'in-scherptevlak' geplaatst is (Garg & Wadhwa, 2018). Om deze extra parameter mee te implementeren in de kwaliteit van de portrait-modus, heeft Google een nieuw algoritme gebouwd op basis van Machine Learning (opnieuw een Convolutional Neural Network).

Hier komen bij de Pixel 4 nog enkele features bij, gaande van een verbetering van de PDAF-technologie (dual-pixel autofocus) tot een verbeterde 'bokeh' (kwaliteit van wazigheid) en de implementatie van een tweede achtercamera. Wadhwa en Zhang (2019) leggen in de Google AI-blog uit hoe de implementatie van deze features ervoor zorgen dat men de portrait modus wellicht zowel van ver als van korte afstand (van het scherp gestelde object) kan toepassen. In de vorige versies van de Google Pixel werd parallax gecreëerd doordat de lens van de camera in tweeën werd gesplitst (PDAF) en doordat er dan tussen de twee perspectieven van een afbeelding 1 millimeter afstand plaatsvindt men via de kleine verschillen de diepte kan meten van de verschillende objecten in de afbeelding. In de Pixel 4 is er echter een tweede achtercamera, waardoor de afstand tussen de twee perspectieven niet langer 1 millimeter bedraagt maar 13 millimeter en de meting van de diepte veel effectiever gebeurt. Toch blijft men in de Pixel 4 gebruik maken van de PDAF-technologie voor o.a. de optimalisatie van de uitlijning van het object dat 'in-focus' staat. De tweede camera is wegens zijn afstand van de eerste namelijk niet in staat om een dieptemeting te maken van de pixels net tegen het object (komen namelijk niet meer in beeld, net zoals het verschil in zicht wanneer we 1 oog sluiten) en voor deze pixels is de PDAF dus essentieel.