

Rapportage Q1 2024

Inleiding

Voor de image restoration taak is een nieuwe aanpak geïmplementeerd, waarbij conventionele diffusion-modellen met een Variational Autoencoder (VAE) en CLIP zijn vervangen door een Denoising Diffusion Probabilistic Model (DDPM). De motivatie voor deze verandering komt voort uit de beperkingen van bestaande pretrained diffusion-modellen, die niet in staat waren om effectief te generaliseren naar domeinen met hoge variabiliteit, zoals onderwaterfotografie. Finetuning van bestaande modellen leverde beperkte verbeteringen op, en de afhankelijkheid van voorgetrainde representaties introduceerde een bias die de modelprestaties beïnvloedde.

Door een DDPM from scratch te trainen, is volledige controle over het trainingsproces en de samenstelling van de trainingsdata mogelijk. Dit biedt flexibiliteit in het toepassen van specifieke noise conditioning en ruisreductietechnieken, die beter aansluiten bij de taak van beeldrestauratie.

Methode

Het ontwikkelde model is gebaseerd op een UNet-architectuur, waarbij een encoder-decoder structuur met residual connections is gebruikt om stabiliteit te waarborgen. De kern van de architectuur bestaat uit een diffusion-proces, waarbij ruis iteratief wordt toegevoegd en vervolgens verwijderd via een reverse diffusion-proces. De DDPM Noise Scheduler maakt gebruik van een lineaire beta-schedule met waarden van $1e-4$ tot 0.02, wat het verloop van ruis over de trainingsstappen bepaalt.

Voor training is een Mean Squared Error (MSE) loss-functie toegepast, waarbij de voorspelde ruis wordt vergeleken met de werkelijke ruis in de noised samples. Als optimizer is Adam gebruikt met een learning rate van $1e-4$. Het model is getraind met een batchgrootte van 32 over 100 epochs.

De trainingsdataset is specifiek samengesteld om te voldoen aan de eisen van onderwaterfotografie. De dataset bevat afbeeldingen met variabele niveaus van degradatie, zoals kleurvervorming, lichtabsorptie en ruis veroorzaakt door zwevende deeltjes. Preprocessing bestaat uit contrastnormalisatie, kleurcorrectie en data augmentation, zodat het model robuuster wordt voor diverse beeldomstandigheden.

Resultaten

De prestaties van het model zijn geëvalueerd met Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) en Learned Perceptual Image Patch Similarity (LPIPS). De gemiddelde PSNR bedraagt 17.47 dB, wat suggereert dat er ruimte is voor verbetering in ruisonderdrukking en detailbehoud. De SSIM-score van 0.5578 duidt op een matige structurele overeenkomst met de grondwaarheidsafbeeldingen, terwijl de LPIPS-score van 0.4221 aangeeft dat de perceptuele kwaliteit inconsistent is en nog verbeterd moet worden.

Bij analyse van individuele beelden blijkt dat PSNR en SSIM sterk variëren afhankelijk van de mate van vervorming in de oorspronkelijke afbeelding. Beelden met zware degradatie worden minder effectief gereconstrueerd, terwijl lichtere vervormingen beter worden hersteld. LPIPS-

scores laten zien dat het model moeite heeft met het correct herstellen van kleurinformatie, wat resulteert in kleurverschuivingen en artefacten in de gereconstrueerde beelden.

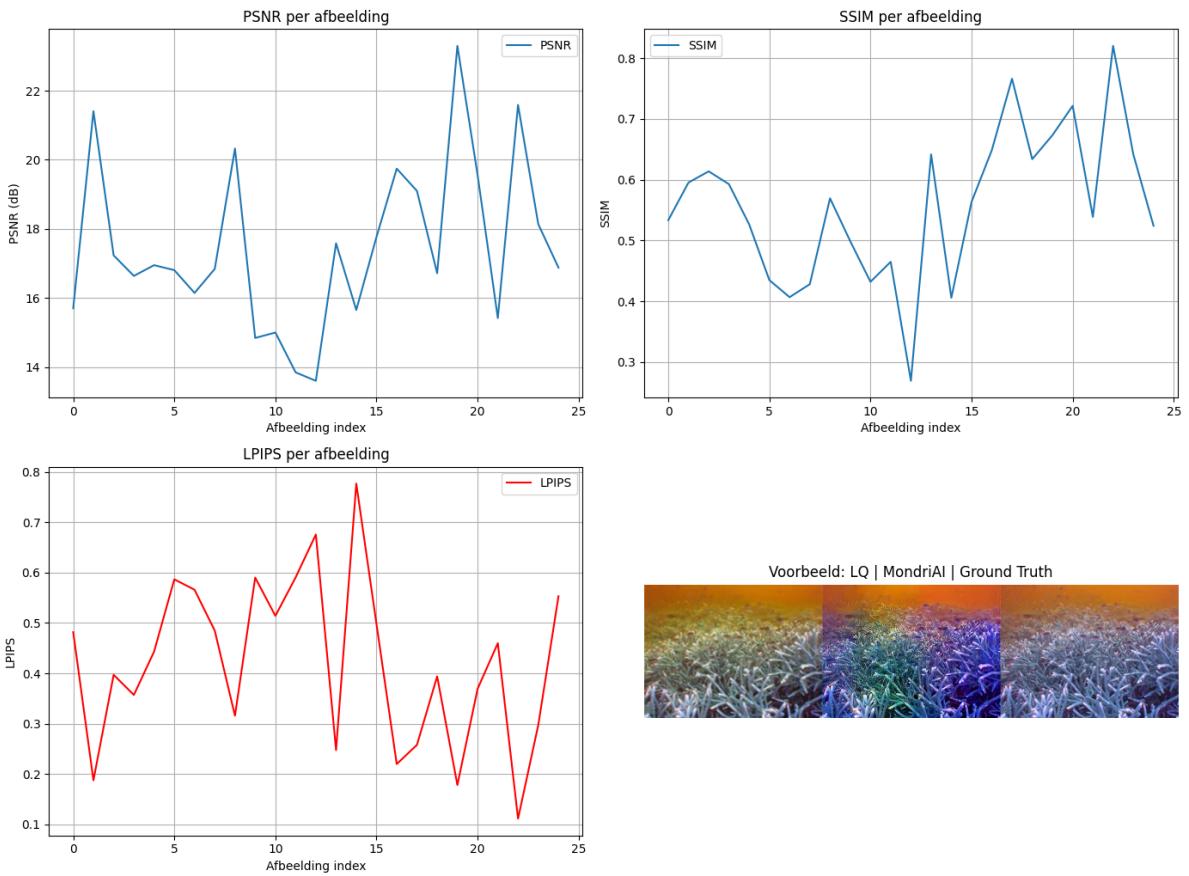
De inferentietijd per afbeelding bedraagt gemiddeld 5.7781 seconden, wat aantoont dat de rekenintensiviteit van het model een knelpunt vormt. Dit is met name een probleem voor toepassingen waar real-time verwerking vereist is.

Discussie

Hoewel het model in staat is om enige vorm van beeldrestauratie uit te voeren, zijn er meerdere aspecten die verbetering behoeven. De huidige dataset kan uitgebreid worden met diverse beelden en complexere degradatiepatronen, zodat het model beter generaliseert naar verschillende onderwateromstandigheden. Daarnaast kan de architectuur geoptimaliseerd worden door gebruik te maken van alternatieve loss-functies, zoals een perceptual loss gebaseerd op deep feature extractie, om de visuele kwaliteit van de gereconstrueerde beelden te verbeteren.

Conclusie

De overstep naar een DDPM-model heeft geleid tot meer controle over het trainingsproces en een betere afstemming op de eisen van onderwaterbeeldrestauratie. De huidige resultaten tonen aan dat het model structurele en perceptuele verbeteringen kan aanbrengen, maar dat er nog aanzienlijke ruimte is voor optimalisatie.



Gemiddelde resultaten voor Val_Dataset:
 PSNR: 17.47 dB
 SSIM: 0.5578
 LPIPS: 0.4221
 Gemiddelde testtijd per afbeelding: 5.7781 seconden

Afbeelding	PSNR (dB)	SSIM	LPIPS
Afbeelding 1	15.70	0.5334	0.4817
Afbeelding 2	21.41	0.5954	0.3876
Afbeelding 3	17.23	0.6138	0.3970
Afbeelding 4	16.64	0.5927	0.3571
Afbeelding 5	16.95	0.5266	0.4432
Afbeelding 6	16.81	0.4349	0.5866
Afbeelding 7	16.15	0.4067	0.5660
Afbeelding 8	16.84	0.4281	0.4848
Afbeelding 9	20.33	0.5694	0.3159
Afbeelding 10	14.84	0.4989	0.5899
Afbeelding 11	15.00	0.4320	0.5141
Afbeelding 12	13.85	0.4649	0.5904
Afbeelding 13	13.60	0.2690	0.6756
Afbeelding 14	17.58	0.6418	0.2476
Afbeelding 15	15.65	0.4057	0.7767
Afbeelding 16	17.77	0.5636	0.4976
Afbeelding 17	19.74	0.6485	0.2199
Afbeelding 18	19.10	0.7663	0.2579
Afbeelding 19	16.72	0.6339	0.3939
Afbeelding 20	23.30	0.6733	0.1783
Afbeelding 21	19.56	0.7216	0.3693
Afbeelding 22	15.42	0.5388	0.4599
Afbeelding 23	21.59	0.8204	0.1116
Afbeelding 24	18.14	0.6417	0.2972
Afbeelding 25	16.88	0.5242	0.5530

Val_Dataset MondriAI underwater test set 2.1 - Voorbeeld 1: Low quality | MondriAI | Ground Truth



Val_Dataset MondriAI underwater test set 2.1 - Voorbeeld 2: Low quality | MondriAI | Ground Truth



Val_Dataset MondriAI underwater test set 2.1 - Voorbeeld 3: Low quality | MondriAI | Ground Truth

