



Universiteit
Leiden

The Netherlands

Using deep learning models in image processing

Xiong, Z.

Citation

Xiong, Z. (2026, May 19). *Using deep learning models in image processing*.
Retrieved from <https://hdl.handle.net/1887/4303646>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4303646>

Note: To cite this publication please use the final published version (if applicable).

Nederlandse samenvatting

Beeldbewerking heeft met de komst van Deep Learning een revolutie doorgemaakt. Dit heeft, in het bijzonder, opmerkelijke voortgang gegeven in het onderzoeksveld variërend van autonome besturing van voertuigen tot medische diagnostiek. Ondanks het succes van de rekenmodellen onderliggend aan Deep Learning, heeft een steeds meer toenemende complexiteit de urgente vragen opgeworpen over hoe deze modellen werken, waarom ze soms niet goed werken, en of deze modellen vertrouwd kunnen worden in kritische toepassingen. In dit proefschrift worden juist deze vragen geadresseerd door het presenteren van een overzicht van de ontwikkelingen in deep-learning in het beeldbewerkingsdomein en daarnaast worden bijdragen geleverd waarin de vooruitgang in model interpretatie en betrouwbaarheid wordt behandeld. Het onderzoek beschreven in dit proefschrift begint vanuit het perspectief van een breed overzicht gevolgd door specifieke methodologische verkenningen. Tezamen vormen de Hoofdstukken de basis voor transparante en robuuste systemen voor computer visie.

In **Hoofdstuk 1** behandelen we de begrippen die van belang zijn voor dit proefschrift. Vervolgens introduceren we de onderzoeksvragen die de basis zijn voor de inhoud van de Hoofdstukken 2 tot en met 5.

In **Hoofdstuk 2** wordt een uitgebreid literatuuroverzicht gepresenteerd over Deep Learning modellen die het werkveld van beeldbewerking domineren. In dit overzicht zijn het uitgangspunt de wiskundige principes die worden gedeeld door convolutionele neurale netwerken en vision-transformers, daarbij beoordelen we hoe deze modellen in parallel zijn geëvolueerd en waar eventuele convergentie wordt gevonden.

In **Hoofdstuk 3** bouwen we verder op het fundament van het voorgaande hoofdstuk en behandelen we interpreteerbaarheid van een kerntaak in computer visie, i.e. semantische segmentatie. We leggen de nadruk op het mechanisme van cross-attentie dat we herformuleren als een probabilistisch model en laten zien dat deze aanpak kan worden

geïnterpreteerd als een k-means clustering in de representatie van de kernmerkruimte. Dit nieuwe perspectief verheldert de veel gebruikte cross-attentie component en het voorziet tevens in een solide basis voor begrip van het gedrag van deze component onder de zogenaamde conditie van volledige supervisie.

In **Hoofdstuk 4** vervolgen we ons onderzoek en bestuderen we het uitdagende “weakly supervised” segmentatie waarbij annotaties op basis van gelabelde pixels niet beschikbaar zijn. We analyseren hoe in afwezigheid van gedetailleerde labels de kenmerkruimte tot stand komt en introduceren een probabilistisch kader waarmee missende achtergrondinformatie door contrastief leren kan worden gecorrigeerd. In samenhang laten de **Hoofdstukken 3 en 4** zien dat een geïntegreerde probabilistische weergave attentiegebaseerde modellen kan verklaren en verbeteren onder verschillende regimes van supervisie. Hierdoor wordt de interpreteerbaarheid van deze modellen aanzienlijk vergroot.

In **Hoofdstuk 5** gebruiken we onze technische bijdrages in een veeleisende klinische toepassing: de analyse van microscoopbeelden van nier-biopten. De uitwerking van onze technische bijdrages resulteert in een betrouwbare en robuuste aanpak. In transplantatiegeneeskunde is het monitoren van een getransplanteerde nier cruciaal, de status van de nier wordt aan de hand van de micro-anatomie in een biopt bestudeerd. De anatomische structuren zitten dicht opeengepakt, vertonen een grote variatie in vorm en grootte en zitten vaak aaneengeplakt. Om een automatische analyse mogelijk te maken introduceren we een generatief model van instantie segmentatie waarbij diffusie modellen worden gecombineerd met vision-transformers. De architectuur van onze methode is expliciet ontworpen om onbalans in de klassen te beperken, objecten te herkennen bij wisselende vergroting en desambiguering van aaneengeplakte structuren door het benadrukken van randinformatie. Door extensieve experimenten kunnen we de robuustheid van onze methode vaststellen en illustreren we de potentie voor het inzetten in de klinische praktijk.

In **Hoofdstuk 6**, tenslotte, vatten we de resultaten en conclusies van ons onderzoek samen en beantwoorden we de onderzoeksvragen. Met de resultaten als basis schetsen we een perspectief voor toekomstige ontwikkelingen in het veld.

Alles bij elkaar biedt dit proefschrift een reflectie over hoe een macroscopisch begrip van de evolutie van een model (**Hoofdstuk 2**) informatie geeft over specifiek onderzoek in interpreteerbaarheid (**Hoofdstuk 3-4**) en betrouwbaarheid (**Hoofdstuk 5**). Daarmee laat het zien hoe onze bijdragen tezamen vooruitgang bieden aan het doel

computer vision systemen te bouwen die krachtig zijn en tegelijkertijd ook begrijpelijk en betrouwbaar. Door het overbruggen van theoretische fundamenteen naar praktische oplossingen, tracht dit proefschrift de weg te bereiden door het implementeren van deep-learning modellen die betrouwbaar zijn onder operationele omstandigheden, waar transparantie en afhankelijkheid even belangrijk zijn als nauwkeurigheid.