



Universiteit  
Leiden  
The Netherlands

## Exploring images with deep learning for classification, retrieval and synthesis

Liu, Y.

### Citation

Liu, Y. (2018, October 24). *Exploring images with deep learning for classification, retrieval and synthesis*. *ASCI dissertation series*. Retrieved from <https://hdl.handle.net/1887/66480>

Version: Not Applicable (or Unknown)

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

Downloaded from: <https://hdl.handle.net/1887/66480>

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

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/66480> holds various files of this Leiden University dissertation.

**Author:** Liu, Y.

**Title:** Exploring images with deep learning for classification, retrieval and synthesis

**Issue Date:** 2018-10-24

# Nederlandse Samenvatting

Een van de belangrijkste doelen van het onderzoek op het gebied van Computer Vision is om machines toe te rusten met het vermogen om visuele data te begrijpen. Gedreven door dit doel presenteren we in dit proefschrift onderzoek op het gebied van de exploratie en analyse van beelden binnen de drie thema's: classificatie, retrieval en synthese.

Ons eerste thema concentreert zich op classificatie op beeld- en pixel-niveau. Ten eerste introduceren we een efficiënt convolutioneel fusie netwerk dat in staat is om adaptieve gewichten aan te leren door middel van de fusie van verschillende tussenliggende lagen van het netwerk waarbij slechts een klein aantal nieuwe parameters benodigd zijn. Daarenboven kan ons neurale netwerk eenvoudig uitgebreid worden om classificaties op pixel-niveau, zoals semantische segmentatie en edge-detectie, uit te voeren. Ons werk is een sterke aanwijzing dat diepe fusie netwerken superieur zijn aan gewone convolutionele netwerken. Ten tweede bestuderen we de classificatie-taak op pixel-niveau voor edge-detectie. In tegenstelling tot eerdere studies welke een supervisie gebruikten die constant blijft voor alle tussenliggende lagen, ontwikkelen wij een diverse supervisie welke zich kan aanpassen aan de diversiteit van de verschillende lagen van het netwerk. Onze methode is daarbij in staat om de diversiteiten in het supervisie-sigitaal in te lijven.

Het tweede thema van dit proefschrift omvat image retrieval en cross-modal retrieval. We construeren een DeepIndex-framework door het opnemen van deep visual features in het geïnverteerde index-schema. We kunnen dan een meervoudig DeepIndex-framework gebruiken om verschillende deep features op het index niveau te integreren. Verder ontwikkelen we een deep matching netwerk voor de unificatie van visuele en tekstuele features ten behoeve van cross-modal retrieval. De bouwsteen van ons netwerk integreert het recurrente mechanisme, het residuale leren en een fusie-module. Deze integratie kan helpen bij de ontwikkeling van feature embeddings terwijl gedeelde parameters behouden blijven. We introduceren cycle-consistent embeddings welke bij het matchen van visuele en tekstuele representaties zowel intermodale correlaties als intra-modale consistentie kan behouden. Voor robuuste afleidingen gebruiken we twee late-fusion benaderingen om matching scores van verschillende embedding features te integreren. Ten slotte exploiteren we een unified netwerk voor zowel matching als classificatie, dit in tegenstelling tot netwerken voor

ofwel multimodale matching of multimodale classificatie. De matching component is in staat om het modaliteits-gat tussen beeld en taal te overbruggen terwijl tegelijkertijd de classificatie component visuele en tekstuele embedding features combineert tot een multimodale representatie.

In ons derde thema bestuderen we twee beeldsynthese applicaties. Ten eerste breiden we cycle-consistente generatieve adversarial netwerken uit voor beeld-naar-beeld vertalingen. Onze uitgebreide modellen maken gebruik van meer generatoren en inwendige cycli om de constraints te versterken terwijl de unsupervised vertaling tussen verschillende beelddomeinen wordt uitgevoerd. Ten tweede introduceren we een nieuw generatief kader om het probleem van persoon-persoon fashion style transfer aan te pakken. Het omvat drie generatieve netwerken welke in een cascade een multi-stage paradigma vormen. Ons kader kan end-to-end getraind worden om de kleren van een persoon te verwisselen met behoud van pose en lichaamsvorm.

We hebben talloze experimenten uitgevoerd om de effectiviteit van onze voorgestelde methoden binnen de drie onderzoeksthema's te verifiëren. Onze resultaten laten veelbelovende verbeteringen zien in vergelijking met verschillende baseline methoden en zijn vergelijkbaar met de beste resultaten uit de research community. Door het uitvoeren van een wijd scala aan taken en applicaties uit het onderzoeksveld geeft dit proefschrift nieuwe contributies, inzichten en bevindingen.