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

English Summary

A core mission of computer vision research is endowing machines with the ability to understand visual data. Driven by it, in this thesis we present research on exploring and analyzing images from three themes: classification, retrieval and synthesis.

Our first theme focuses on image-level and pixel-level classification. Firstly, we propose an efficient convolutional fusion network that can learn adaptive weights by fusing different intermediate layers with adding only a few parameters. In addition, our proposed neural network can be extended for pixel-level classification such as semantic segmentation and edge detection. Our work suggests the superiority of deep fusion networks over plain convolutional neural networks. Secondly, we further study the pixel-level classification task for edge detection. In contrast to prior works that use a fixed supervision for all intermediate layers, we develop diverse supervision that can adapt to the diversities of different layers. Our method can incorporate the diversities into the supervisory signals.

The second theme of this thesis includes image retrieval and cross-modal retrieval. We build a DeepIndex framework by incorporating deep visual features into the inverted index scheme. Subsequently, we can leverage a multiple DeepIndex framework to integrate different deep features at an indexing level. Furthermore, we develop a deep matching network to unify visual and textual features for cross-modal retrieval. The building block in our network integrates the recurrent mechanism, the residual learning and a fusion module. This integration can help promote feature embeddings while retaining the shared parameters. We propose cycle-consistent embeddings which can preserve both inter-modal correlations and intra-modal consistency while matching visual and textual representations. For a robust inference, we leverage two late-fusion approaches to integrate the matching scores of different embedding features. Lastly, in contrast to either multi-modal matching or multi-modal classification, we exploit a unified network for joint matching and classification. The matching component can bridge the modality gap between vision and language, and simultaneously the classification component is used to combine visual and textual embedding features to be a multi-modal representation.

Our third theme studies two applications about image synthesis. Firstly, we extend cycle-consistent generative adversarial networks for image-to-image translation. Our extended models make use of more generators and inner cycles to enhance the

constrains while performing the unsupervised translation between different image domains. Secondly, we propose a novel generative framework for addressing the problem of person-to-person fashion style transfer. It includes three generative networks that are cascaded in a multi-stage paradigm. Our framework can be trained end-to-end to swap the clothes of person images while preserving their pose and body shape.

We have conducted numerous experiments to verify the effectiveness of our proposed methods for the three research themes. Our results demonstrate promising improvements over various baseline methods, and are comparable with the state-of-the-art results from the research community. By performing a wide range of tasks and applications in the field, this thesis provides novel contributions, insights and findings.