

DISS. ETH NO. 23595

# **Towards Cost-Effective and Performance-Aware Vision Algorithms**

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Science (Dr. sc. ETH Zürich)

presented by

**Dengxin Dai**

Master of Science

born November 15, 1986

citizen of China

accepted on the recommendation of

Prof. Dr. Luc Van Gool, examiner

Prof. Dr. Gerhard Schmitt , co-examiner

Prof. Dr. Bernt Schiele, co-examiner

2016

# Abstract

Computer vision has leaped forward during the last decade, and now is able to recognize objects of thousands of categories and reconstruct 3D scenes at city- or world-scale. However, the field still has to find means to keep up with the exploration of the massive amounts of data being captured on a daily basis. This is mainly due to the lack of sufficient training annotations and the lack of computational resources. The thesis is dedicated to mitigate the problem.

Firstly, we elaborate two strategies to reduce the annotation costs in order to train vision algorithms: (1) developing smart annotation approaches for efficient, large-scale annotations; and (2) learning better feature representations using unlabeled data; We develop algorithms for the strategies and show - in the context of recognition tasks - that they are able to considerably reduce the annotation costs for the training of recognition algorithms.

Secondly, in addition to reducing annotation cost, we also examine how to reduce the computational cost associated with the training and testing of recognition algorithms. This research has lead to two contributions: (1) two efficient solvers for linear and kernel SVM+, significant speeding up the training process of SVM+ to explore privileged information; and (2) a method to allow computationally cheap features to imitate alternative features that perform better but are computationally more expensive. The imitation significantly improves the performance of the cheap features while retaining their efficiency.

Thirdly, as images keep growing in size, vision algorithms need to be more intelligent and self-aware of their performance. To this aim, we have developed approaches to predict and evaluate the performance of vision algorithms at two levels of granularity: *Failure Prediction* and *Failure Diagnosis*. The two methods are evaluated on texture synthesis and image segmentation respectively, and their potentials in reducing computational time are examined as well.

# Zusammenfassung

Die Bilderkennung hat in den letzten zehn Jahren einen enormen Entwicklungsschub hinter sich und ist nun in der Lage, tausende von Objektkategorien zu erkennen und 3D-Szenen zu rekonstruieren, deren Komplexität von der Stadt- bis zur Welt-Größenordnung reicht. Allerdings bleibt noch zu erforschen, wie wir mit den massiven Datenmengen, die täglich erfasst werden, umgehen sollen. Uns mangelt es vor allem an ausreichenden annotierten Trainingsdaten und Rechenressourcen. Dort liegt der Fokus dieser Arbeit. Erstens erarbeiten wir zwei Strategien, um die Annotationskosten zu reduzieren, die wir fürs Training von Bilderkennungsalgorithmen benötigen: (1) Entwicklung intelligenter Annotationstechniken, die uns ermöglichen effizient eine Grosszahl von Beispielen zu annotieren; und (2) Techniken zum automatischen Erlernen von Darstellungsformen mit Hilfe von unmarkierten Daten. Wir entwickeln Algorithmen für diese Strategien und zeigen (im Rahmen von Erkennungsaufgaben), dass sie die Annotationskosten für das Training von Bilderkennungssystemen erheblich reduzieren können. Zweitens wird neben der Reduzierung der Annotationskosten auch untersucht, wie die mit dem Training und dem Testen verbundenen Rechenkosten verringert werden können. Unsere Studie erbrachte in diesem Feld die folgenden zwei Beiträge: (1) zwei effiziente Lösungsverfahren für lineare und Kern-SVM+, die das Training von SVM+ beschleunigen und dazu bewegen, die privilegierten Informationen zu untersuchen; und (2) ein Verfahren zum Erlernen von Darstellungsformen, welche deutlich schneller als die anderen Darstellungsformen berechnet werden. Diese effizienten Darstellungsformen imitieren die reicherer, rechnerisch aufwendigeren Darstellungsformen, weisen dabei aber eine ähnlich gute Leistung auf. Des weiteren müssen die Bilderkennungsalgorithmen intelligenter und sich über ihre eigene Leistung bewusster werden, da Bild-Datensätze immer grösser werden. Zu diesem Zweck entwickelten wir Ansätze, um die Leistung von Algorithmen auf zwei verschiedenen Abstraktionsstufen vorherzusagen und zu evaluieren: Fehlervorhersage und Fehlerdiagnose. Wir evaluieren diese zwei Methoden im Kontext von Textur-Synthetisierung und Bildsegmentierung und untersuchen, wie man diese weiterhin beschleunigen kann.