

Lernen mit wenigen Beispielen für die visuelle Objekterkennung

Erik Rodner
Lehrstuhl Digitale Bildverarbeitung
Friedrich-Schiller Universität Jena
Erik.Rodner@uni-jena.de

Abstract: Das maschinelle Lernen aus wenigen Beispielen ist ein wichtiges und entscheidendes Problem bei vielen visuellen Erkennungsaufgaben, besonders in industriellen Anwendungen. Im Gegensatz zum Menschen benötigen viele aktuelle Verfahren meistens Hunderte von beschrifteten Beispielbildern. Die Dissertation "*Learning with Few Examples for Visual Recognition Problems*" beschäftigt sich mit diesem Problem und stellt Lösungsmöglichkeiten vor, welche sich auf die Verwendung zweier Konzepte stützen: *Lerntransfer* und *Ein-Klassen-Klassifikation*. Das folgende Dokument bietet eine Zusammenfassung der Ergebnisse der Dissertation.

1 Einleitung

Die Dissertation beschäftigt sich mit Verfahren der visuellen Objekterkennung, welche das Ziel verfolgen, automatisch semantische Informationen aus Bildern zu extrahieren. Dabei sollen zum Beispiel Objekte bekannter Kategorien in einem Bild erkannt und lokalisiert werden. Weiterhin soll die Maschine die Erscheinungsformen einer Objektkategorie selbstständig aus beschrifteten Beispielbildern lernen. Die Fähigkeit, diese Art der automatischen Bildanalyse durchzuführen, ist sowohl in der Robotik als auch bei zahlreichen Anwendungen zwingend notwendig. In den letzten Jahren lässt sich ein drastischer Anstieg an komplexen industriellen Problemstellungen verzeichnen, welche ohne Verfahren des maschinellen Lernens nicht realisierbar sind. Als prägnantes Beispiel sei hier die Fußgängerdetektion [DWSP11] und zahlreiche andere Fahrerassistenzsysteme aufgeführt. Ein Hauptproblem ist die Verfügbarkeit von repräsentativen Lernbeispielen, da die Beschriftung bei vielen Anwendungen zeit- und kostenintensiv ist. *Ziel der Dissertation ist es daher die Anzahl der notwendigen Lernbeispiele durch spezielle Verfahren des maschinellen Lernens zu reduzieren.*

In Abbildung 1 sind die drei Hauptabstraktionsebenen der visuellen Objekterkennung dargestellt. Diese richten sich nach der Art der gewünschten Ausgabe des Systems und des Detaillierungsgrades. Während der Dissertation wurden alle drei Bereiche betrachtet [RD10, FRD10]. Die entwickelten Verfahren sind allgemein für viele Aufgaben des maschinellen Lernens geeignet.

Die vorliegende Arbeit ist wie folgt aufgebaut: zunächst wird allgemein auf die Schwierigkeiten der visuellen Objekterkennung und des Lernens aus wenigen Beispielen einge-

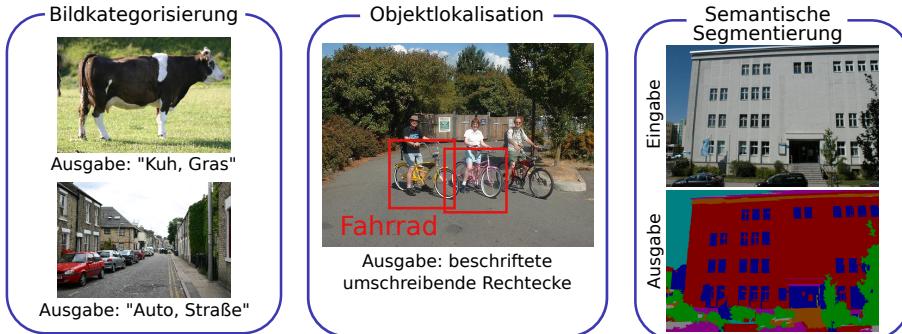


Abbildung 1: Unterschiedliche Aufgabenstellungen der visuellen Objekterkennung: Bildkategorisierung [RD10] (Beschriftung des gesamten Bildes), Objektlokalisation (Beschriftete umschreibende Rechtecke) und Semantische Segmentierung [FRD10] (Beschriftung jedes einzelnen Bildpixels).

gangen. Danach wird das Konzept des Lerntransfers und die in der Dissertation entwickelten Verfahren skizziert. Abschnitt 5 liefert einen Überblick über die Algorithmen der Ein-Klassen-Klassifikation, welche den zweiten Schwerpunkt der Dissertation darstellen. Realisierte Anwendungen werden in Abschnitt 6 kurz zusammengefasst. Abschließend folgt eine Zusammenfassung der Resultate.

Die Zusammenfassung beschreibt nur die Aspekte der Klassifikation. Für eine Beschreibung der Merkmalsauswahl für einzelne Anwendungen sei auf die Dissertation verwiesen.

2 Herausforderungen bei der visuellen Objekterkennung

Bei den Beispielbildern in Abbildung 1 lassen sich die Schwierigkeiten und die Komplexität des automatischen Lernens von Objektkategorien gut erkennen und folgendermaßen zusammenfassen:

1. Die Erscheinungen einzelner Objektkategorien variieren sehr stark durch unterschiedliche Rotationen, Skalierungen, andere Perspektiven, nicht-starre Deformationen, farbliche Gestaltung, Unterkategorien anderer Ausprägung (z.B. verschiedene Arten von Vegetation in den Bildern auf der rechten Seite).
2. Bestimmte Kategorien sind ähnlich zueinander und lassen sich schwierig voneinander trennen (z.B. Kategorie Fenster und Tür auf der rechten Seite).
3. Objekte können sich gegenseitig verdecken (zu sehen in der Straßenszene unten links) und 3D-Informationen sind bei Einzelaufnahmen nicht direkt verfügbar.
4. Die Darstellung von Kategorien und Bildelementen, welche nicht erkannt werden sollen, erschwert die Erkennung zusätzlich (z.B. die Fahrradfahrer im Bild in der Mitte).

Vor allem die große Variabilität der Objekte einer Kategorie lässt sich meist nur durch die Angabe vieler beschrifteter und repräsentativer Beispiele erlernen. So werden zum Beispiel zum Anlernen von Fußgängerdetektoren oft mehrere tausende Beispiele benötigt [DWSP11].

Das Problem bei wenigen Lerndaten manifestiert sich auch als schlecht gestelltes Optimierungsproblem, welches im Folgenden näher erläutert werden soll: Das Lernen von visuellen Aufgaben kann mathematisch als das Schätzen einer Abbildung $f : \mathcal{X} \rightarrow \mathcal{Y}$ von der Menge \mathcal{X} aller Bilder in die Menge \mathcal{Y} aller möglichen Beschriftungen angesehen werden. Die Funktion f wird im Falle der Klassifikation (\mathcal{Y} ist diskret, z.B. $\mathcal{Y} \in \{-1, 1\}$) als Entscheidungsfunktion bezeichnet. Die Schätzung oder das Lernen basiert dabei auf einem gegebenen Lerndatensatz \mathcal{D} , welcher n Bilder $x_i \in \mathcal{X}$ und deren Beschriftungen $y_i \in \mathcal{Y}$ enthält. Wird die Aufgabe als reines Schätzproblem betrachtet, führt dies unmittelbar zum entscheidenden Dilemma der Objekterkennung: auf der einen Seite ist die Menge aller möglichen Funktionen f und der Eingaberaum \mathcal{X} selbst hochdimensional, auf der anderen Seite existieren nur wenige gegebene Datenpunkte. Ohne weitere Zusatzinformationen ist diese Situation vergleichbar mit der Regression einer komplizierten Funktion (z.B. Polynom hohen Grades) mit einer geringen Anzahl von Abtastwerten. Genau die Einbindung von zusätzlichem Wissen durch einen sogenannten Lerntransfer (englisch: *knowledge transfer* oder *transfer learning*) ist das Schlüsselkonzept, welches das Lernen aus wenigen Beispielen ermöglicht.

3 Lerntransfer

Betrachtet man die menschlichen Erkennungsleistungen, so ist es anscheinend trotz der beschriebenen Schwierigkeiten beim Lernen visueller Aufgaben als Mensch möglich, neue Objektkategorien oft mit nur einem Beispiel robust zu erlernen [Bie87]. Welche Zusatzinformationen werden aber vom menschlichen Erkennungssystem ausgenutzt um dies zu ermöglichen? Als häufiger Punkt wird die automatische Verwendung von Vorwissen ähnlicher Aufgabenstellungen für das Erlernen einer neuen Aufgabe angeführt. Intuitiv veranschaulicht dies der Lerntransfer beim Erlernen von Sprachen: der Aufwand, eine neue Sprache zu erlernen, ist erheblich geringer, wenn schon verwandte und ähnliche Sprachen bekannt sind, z.B. Französisch und Spanisch. In Abbildung 2 wird dies für die visuelle Objekterkennung anhand der Bildkategorisierung mit Tierklassen illustriert. Bei diesen Tierkategorien existieren viele Gemeinsamkeiten, wie etwa ähnliche Texturmerkmale (Zebra) oder eine ähnliche Konstellation von Objektteilen (Nashorn, Zebra). Die Variation dieser visuellen Komponenten kann daher robust von den verwandten Klassifikationsaufgaben erlernt werden.

In der Dissertation werden mehrere Verfahren vorgestellt, die entwickelt wurden um dieses Konzept beim maschinellen Lernen umzusetzen. Dabei werden unterschiedliche Wissensrepräsentationen von einer Klassifikationsaufgabe (*Unterstützungsaufgabe*) auf eine neue Aufgabe (*Zielaufgabe*) übertragen.

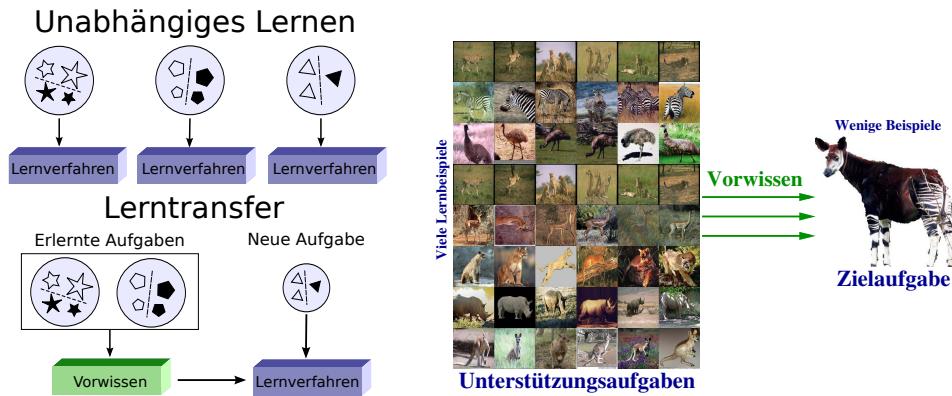


Abbildung 2: (Links) Schematischer Vergleich zwischen unabhängigen Lernen und *Lerntransfer*, (Rechts) Darstellung des Konzeptes des Lerntransfers bei der visuellen Objekterkennung: Durch gezielte Ausnutzung der Ähnlichkeit der neuen Kategorie *Okapi* zu bekannten Objektkategorien mit vielen Lernbeispielen ist eine Reduzierung der Anzahl der notwendigen Lernbeispiele möglich.

4 Adaptiver Lerntransfer mit Gauß-Prozess-Klassifikatoren

Im Rahmen der Dissertation wurde ein Verfahren entwickelt welches auf Kernfunktionen basiert und daher einen nicht-parametrischen Wissenstransfer ermöglicht. Ein besonderer Vorteil dieser Methode ist es, Klassifikationsaufgaben, von denen Wissen transferiert werden soll, automatisch auszuwählen und den Einfluss des Transfers zu adaptieren. Dies wird durch eine effiziente Modellselektion und der Verwendung von semantischen Ähnlichkeiten zwischen Kategoriebegriffen ermöglicht (siehe Übersicht in Abbildung 3). Zunächst wird ein kurzer Überblick über die Gauß-Prozess-Regression und Klassifikation gegeben, da diese ein methodisches Kernelement späterer Algorithmen ist.

Gauß-Prozess-Regression und Klassifikation Viele Klassifikationsverfahren basieren auf einer Parametrisierung $f(\mathbf{x}; \boldsymbol{\theta})$ der Entscheidungsfunktion. Ausgehend von den Lerndaten \mathcal{D} wird ein Parameter $\boldsymbol{\theta}$ bestimmt, welcher die A-posteriori-Wahrscheinlichkeit $p(\boldsymbol{\theta} | \mathcal{D})$ maximiert (vgl. MAP-Schätzung). Die Gauß-Prozess-Regression und Klassifikation kann hingegen anders motiviert werden. Grundidee ist die Betrachtung der Funktion f direkt als Zufallsvariable und die Annahme, dass f gemäß eines Gauß-Prozesses (GP) verteilt ist, d.h. $f \sim \mathcal{GP}(0, \mathcal{K})$. Die Funktion $\mathcal{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{Y}$ ist die Kovarianzfunktion des Gauß-Prozesses und modelliert die Korrelation $\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')]$ von zwei Ausgaben anhand der Ähnlichkeit von \mathbf{x} und \mathbf{x}' .

Im Kontext des maschinellen Lernens wird \mathcal{K} oft als Kern(el)funktion bezeichnet und es lassen sich etliche Formen dieser Funktion zur Modellierung heranziehen. Beispielhaft sei hier die Gaußkernfunktion angeführt:

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \cdot \|\mathbf{x} - \mathbf{x}'\|^2) . \quad (1)$$

An dieser Funktion lässt sich gut erkennen, dass Beispiele mit einer geringen Distanz im Eingaberaum auch zu einer hohen Korrelation der entsprechenden Funktionswerte führen.

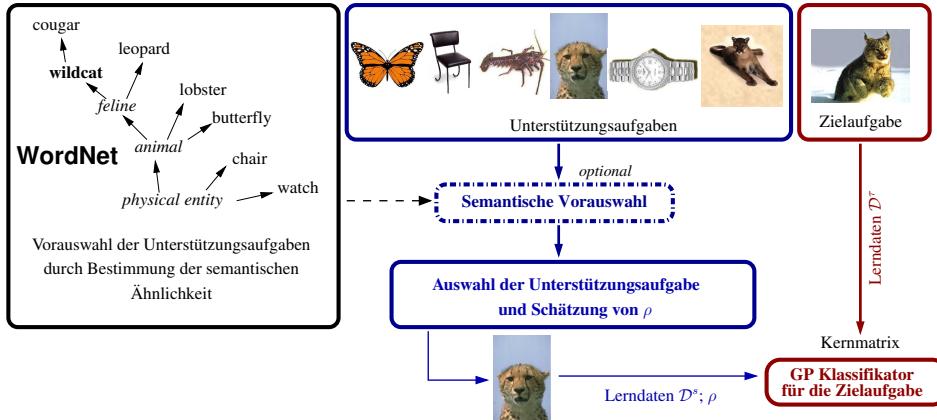


Abbildung 3: Schema des adaptiven Lerntransfers mit Gauß-Prozessen: Aus einer Menge von Klassifizierungsaufgaben wird durch einen zweistufigen Prozess eine Unterstützungsaufgabe ausgewählt um das Lernen einer Zielaufgabe anzureichern. Die Auswahl erfolgt auf der Basis von semantischen Ähnlichkeiten und visuellen Informationen.

Tatsächlich ist dies eine der notwendigen Hauptannahmen des maschinellen Lernens: ähnliche Eingaben sollten zu ähnlichen Ausgaben führen. Mit weiteren Annahmen kann die A-posteriori-Verteilung $p(y_* | \mathbf{x}_*, \mathcal{D})$ der Ausgabe y_* eines neuen Beispiels \mathbf{x}_* hergeleitet werden [RW05]. Im Rahmen dieser Zusammenfassung soll auf mathematische Details verzichtet und nur die Gleichung für den Schätzwert angegeben werden:

$$\mathbb{E}(y_* | \mathbf{x}_*, \mathcal{D}) = \mathbf{k}_*^T (\mathbf{K} + \sigma_\varepsilon^2 \cdot \mathbf{I})^{-1} \mathbf{y} . \quad (2)$$

Bei diesem Modell wurde angenommen, dass die gegebenen Ausgaben zusätzlich mit additivem, normalverteilten Rauschen $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ gestört sind. Die Ausgaben y_i des Lerndatensatzes sind im Vektor $\mathbf{y} \in \mathcal{Y}^n$ zusammengefasst, $\mathbf{K} \in \mathbb{R}^{n \times n}$ bezeichnet die Kernmatrix, welche die paarweisen Werte der Kernfunktion von den Lerndaten beinhaltet, und im Vektor $\mathbf{k}_* \in \mathbb{R}^n$ sind die Werte der Kernfunktion der Lerndaten mit dem neuen Beispiel \mathbf{x}_* gespeichert. Die Annahme von normalverteilten Rauschen ist natürlich eine sehr restriktive Annahme, gerade bei der Klassifikation mit diskreten Beschriftungen $y \in \{-1, 1\}$. Andere Rauschmodelle führen hingegen zu Schätzgleichungen, welche nicht in geschlossener Form und nur approximativ ermittelt werden können. In den Untersuchungen der Dissertation zeigte sich, dass eine Anwendung der GP Regression auch auf Klassifizierungsaufgaben sinnvoll ist und bei vielen Anwendungen zu besseren Ergebnissen als reine Klassifikationsmodelle führt.

Abhängige Gauß-Prozesse und Lerntransfer Eine entscheidende Frage ist, wie das Konzept des Lerntransfers bei der GP Klassifikation verwendet werden kann. Es sei im Folgenden davon ausgegangen, dass genau zwei binäre Klassifizierungsaufgaben gegeben sind, eine Unterstützungsaufgabe s mit Lerndaten \mathcal{D}^s und eine Zielaufgabe τ mit Lerndaten \mathcal{D}^τ . Anstatt die Klassifikatoren für diese Aufgaben jeweils unabhängig voneinander zu lernen ist es das Ziel ein gemeinsames Lernen zu ermöglichen. Durch diesen Schritt ist ein Transfer von Informationen zwischen den Aufgabenstellungen realisierbar.

Ein entscheidendes Konzept sind sogenannte abhängige Gauß-Prozesse [BCW08] (englisch: *dependent Gaussian processes*). Jeder der Klassifikationsaufgaben ist eine Funktion zugeordnet. Diese seien mit f^s für die Unterstützungsaufgabe und f^τ für die Zielaufgabe bezeichnet. Als grundlegende Annahme des Lerntransfers wurden die Ähnlichkeiten der Klassifikationsaufgaben zueinander vorausgesetzt. Diese Annahme kann nun unmittelbar als Korrelation zwischen den Funktionen modelliert werden und es lässt sich folgende gemeinsame A-priori-Annahme aufstellen mit $j, j' \in \{s, \tau\}$:

$$\mathbb{E}(f^j(\mathbf{x})f^{j'}(\mathbf{x}')) = \mathcal{K}((j, \mathbf{x}), (j', \mathbf{x}')) = \begin{cases} \mathcal{K}^X(\mathbf{x}, \mathbf{x}') & \text{wenn } j = j' \\ \rho \cdot \mathcal{K}^X(\mathbf{x}, \mathbf{x}') & \text{sonst} \end{cases}. \quad (3)$$

Der Parameter ρ gibt die Korrelation der Klassifikationsaufgaben an. Gleichung (3) kann als Erweiterung der Kernfunktion betrachtet werden und erlaubt es daher den Erwartungswert der A-posteriori-Verteilung von y_* direkt aus Gleichung (2) abzuleiten:

$$\begin{aligned} \mathbb{E}(y_* | \mathbf{x}_*, \mathcal{D}^s, \mathcal{D}^\tau) &= \mathbf{k}_*(\rho)^T (\mathbf{K}(\rho) + \sigma_\varepsilon^2 \cdot \mathbf{I})^{-1} \mathbf{y} \\ &= \begin{bmatrix} \mathbf{k}_{\tau*} \\ \rho \mathbf{k}_{s*} \end{bmatrix}^T \left(\begin{pmatrix} \mathbf{K}_{\tau\tau} & \rho \mathbf{K}_{\tau s} \\ \rho \mathbf{K}_{s\tau}^T & \mathbf{K}_{ss} \end{pmatrix} + \sigma_\varepsilon^2 \cdot \mathbf{I} \right)^{-1} \begin{bmatrix} \mathbf{y}_\tau \\ \mathbf{y}_s \end{bmatrix}. \end{aligned} \quad (4)$$

Die Indizierung mit s und τ dient der Zuordnung der Werte in den Vektoren und Matrizen zu Lernbeispielen der Zielaufgabe τ oder der Unterstützungsaufgabe s . Bei einer Wahl des Parameters durch $\rho = 0$ erhalten wir das ursprüngliche unabhängige Lernen der Klassifikatoren und bei $\rho = 1$ werden alle Lernbeispiele von s direkt für τ verwendet. Der Parameter erlaubt daher einen adaptiven Lerntransfer.

Auswahl von Unterstützungsklassen Ein Kernelement des entwickelten Verfahrens ist die automatische Schätzung des Parameters ρ und die Auswahl einer Unterstützungsaufgabe aus mehreren Klassifikationsaufgaben mit vielen Lernbeispielen. Dafür wurde eine effiziente Modellselektion mit Leave-one-out Schätzungen entwickelt [RD10], welche die Unterstützungsaufgabe mit der größten zu erwartenden Klassifikationsleistung auswählt.

Eine Auswahl, die nur auf den Bildinformationen basiert, kann natürlich bei der Verwendung von wenigen Lernbeispielen für die Zielaufgabe auch nachteilig sein und so einem sogenannten *negativem Transfer* führen. Daher ist es ratsam, noch zusätzliche Informationsquellen mit einzubeziehen. So ist es zum Beispiel möglich, linguistische semantische Datenbanken, wie etwa WordNet [DDS⁺09], zu verwenden, um die Ähnlichkeit der Klassifikationsaufgaben auch anhand der Kategoriebezeichnungen durchzuführen. Ein optionaler Schritt des neuen Verfahrens zum adaptiven Lerntransfer ist daher eine Vorauswahl aufgrund von semantischen Ähnlichkeiten, d.h. es werden K Objektkategorien ausgewählt, welche zu der neuen Kategorie am ähnlichsten sind. In den Experimenten konnte gezeigt werden, dass eine Kombination von visuellen und semantischen Informationen vorteilhaft ist, da diese sich ergänzen.

Quantitative Auswertung Alle Methoden wurden quantitativ im Rahmen der Bildkategorisierung ausgewertet. Die Ergebnisse zeigen einen signifikante Steigerung der Erkennungsleistung im Vergleich zu aktuellen Methoden des Lerntransfers und Verfahren, welche keine zusätzlichen Lerndaten anderer Klassifikationsaufgaben verwenden. Abbildung 4 enthält einen Teil der durchgeföhrten Auswertungen, bei denen der Vorteil des

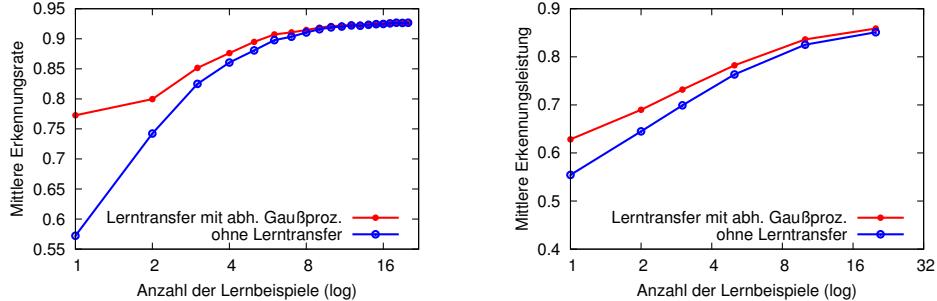


Abbildung 4: Beispielergebnisse der Auswertung des adaptiven Lerntransfers bei der Bildkategorisierung: (Links) mittlere Erkennungsrate bei drei Aufgabenstellungen der Caltech-256 Datenbank, (Rechts) mittlere Erkennungsleistung (average precision) von Aufgabenstellungen der Caltech-101 Datenbank.

Lerntransfers bei wenigen Lernbeispielen gut erkennbar ist. Weiterhin ist zu sehen, dass ab einer gewissen Anzahl von Beispielen, die Informationen im Lerndatensatz der Zielaufgabe genügen und das unabhängige Lernen gleich gute Ergebnisse erzielt wie das Verfahren des Lerntransfers. Für eine ausführliche Beschreibung der Experimente sei auf die Dissertation verwiesen.

Zusammenfassung weiterer Verfahren Eine weitere in der Dissertation vorgestellte Methode erweitert Entscheidungsbaumklassifikatoren um die Möglichkeit, Vorwissen von bereits erlernten Entscheidungsbäumen anderer Aufgaben zu verwenden [RD11]. Zusätzlich wurde eine Ansatz vorgestellt, welcher Informationen über die Merkmalsrelevanz transferiert, um den Lernprozess von randomisierten Entscheidungswäldern anzureichern. Für eine detaillierte Beschreibung sei auf die Dissertation verwiesen.

5 Ein-Klassen-Klassifikation

Eine weitere wichtige Art von Aufgabenstellungen mit wenigen Lernbeispielen sind solche, bei denen nur Lerndaten für eine einzige Klasse vorhanden sind. Dieses Szenario ist besonders häufig bei der Defekt- oder Anomaliedetektion zu finden. So sind zum Beispiel viele Bilder eines fehlerfreien Werkstücks vorhanden, jedoch gibt es nur wenige Aufnahmen von fehlerhaften Elementen. Idee vieler Verfahren der Ein-Klassen-Klassifikation oder Ausreißerdetektion ist es, die Verteilung der fehlerfreien Beispiele zu modellieren (z.B. mit Normalverteilungen). Eine Einschätzung eines neuen Beispiels kann dann aufgrund der Likelihood dieser Verteilung oder allgemein eines Neuheitsmaßes erfolgen. Diese Idee ist im linken Teil von Abbildung 5 noch einmal veranschaulicht.

Zur Lösung von Ein-Klassen-Problemen wurden neue Ansätze in der Dissertation entwickelt und vorgestellt, welche direkt vom Konzept der Regression und Klassifikation mit Gauß-Prozessen abgeleitet wurden. So kann unter Annahme eines mittelwertfreien Gauß-Prozesses als A-priori-Verteilung für die latente Funktion f direkt die Gauß-Prozess-



Abbildung 5: (Links) Veranschaulichung der Ein-Klassen-Klassifikation (Rechts) Anwendung der Ein-Klassen-Klassifikation bei der Schätzung von Attributen.

Regression auf die Ausgaben $y_i = 1$ angewendet werden und es ergibt sich:

$$\mathbb{E}(y_* \mid \mathbf{x}_*, \mathcal{D}) = \mathbf{k}_*^T (\mathbf{K} + \sigma_\varepsilon^2 \cdot \mathbf{I})^{-1} \mathbf{1}. \quad (5)$$

Dieser Erwartungswert ist direkt als Neuheitsmaß einsetzbar. Weitere Maße ergeben sich unter Einbeziehung der Standardabweichung der Schätzung und bei Verwendung von approximativen Inferenzmethoden, auf die an dieser Stelle aber nicht näher eingegangen werden soll.

Die entwickelten Verfahren weisen viele Gemeinsamkeiten zu bekannten Ansätzen wie etwa *support vector data description* [TD04] auf und es lässt sich sogar zeigen, dass Standardverfahren, wie etwa Parzen-Dichteschätzung oder Normalverteilungsklassifikatoren, durch die neuen Verfahren verallgemeinert werden. Ein großer Vorteil ist, dass die Algorithmen zur Klasse der nicht-parametrischen Verfahren gehören, d.h. alle Lerndaten werden bei der Klassifikation eines Beispiels direkt verwendet. In Experimenten wurde in der Dissertation gezeigt, dass die neuen Verfahren zu ähnlichen und sogar oft zu besseren Erkennungsraten als bisherige Methoden führen [KRD10]. Ein weiterer wichtiger Vorteil ist die einfache Implementierung der Algorithmen trotz ihrer theoretischen Komplexität.

Auf der rechten Seite von Abbildung 5 ist die Anwendung der Verfahren für die Schätzung von Attributen zu sehen. Der Klassifikator wurde mit einer speziellen Art von Stuhl (Kategorie: *windsor chair*) angelernt. In der Erkennungsphase ist es dann möglich eine Menge von Bildern nach der vorhandenen Stärke dieses Attributes zu sortieren.

6 Weitere Anwendungen

Im Folgenden werden weitere Anwendungen der Verfahren vorgestellt, welche im Rahmen der Dissertation studiert wurden.

Defektlokalisierung Die Nützlichkeit der Verfahren der Ein-Klassen-Klassifikation wurde anhand der schwierigen Aufgabenstellung der Defektlokalisierung bei Drahtseilen demonstriert. Die Ergebnisse der Experimente zeigen deutlich, dass die vorgestellten Methoden in der Lage sind, bessere Erkennungsergebnisse als bisherige Standardverfahren (z.B. GMM) zu erzielen und Hinweise auf mögliche Defekte zu liefern. Ein Beispielergebnis ist in Abbildung 6 dargestellt.

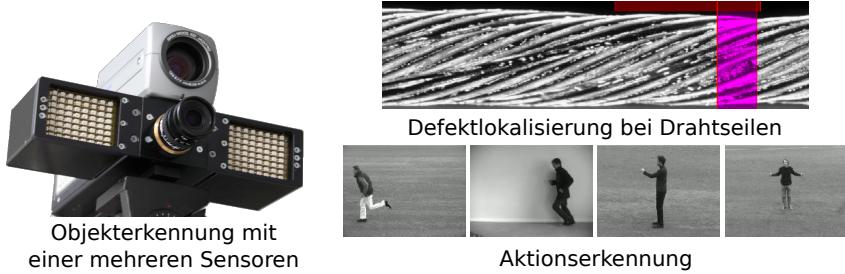


Abbildung 6: Übersicht über die weiteren untersuchten Anwendungen der entwickelten Methoden: Objekterkennung mit Farb- und Time-of-Flight Kameras, Aktionserkennung und -detektion, Defektlokalisierung. Im rechten oberen Bild ist die Erkennung eines Drahtbruches mit dem entwickelten Verfahren zu erkennen. Die automatische Markierung ist in magenta dargestellt und die rote Markierung am oberen Rand zeigt die manuelle Annotation eines Experten.

Aktionserkennung Das Ziel der visuellen Aktionserkennung ist die Erkennung von Aktionen in Videoaufnahmen. Verfahren der Ein-Klassen-Klassifikation können bei dieser Anwendung vorteilhaft sein, da keine Modellierung von Negativbeispielen, d.h. Sequenzen ohne eine Aktion der Kategorie, notwendig ist. Bei den quantitativen Untersuchungen zeigte sich, dass die Ergebnisse stark von der Wahl des Hyperparameters der Kernfunktion abhängen, aber grundsätzlich eine Detektion von Aktionen möglich ist.

Generische Objekterkennung mit mehreren Sensoren Ein zusätzlicher Aspekt, welcher in der Dissertation untersucht wurde, ist die Entwicklung eines Systems zur generischen Objekterkennung, welches die Sensorinformationen einer Farb- und einer Time-of-Flight-Kamera kombiniert. Eine Time-of-Flight Kamera liefert, ähnlich zur aktuell üblichen Kinect-Kamera der Firma PrimeSense, Tiefendaten in Echtzeit. Dadurch können 3D-Informationen gewonnen werden, die bei manchen Erkennungsaufgaben entscheidend sind. In der Dissertation wurde untersucht, wie eine optimale Fusion der Sensordaten für eine bestimmte Klassifikationsaufgabe erfolgen kann. Dabei wurde erneut ein GP-Klassifikator eingesetzt, welcher es ermöglicht mehrere Kernfunktionen linear gewichtet zu kombinieren. In Experimenten zeigte sich, dass diese Kombination besonders bei wenigen Lerndaten vorteilhaft ist. Insgesamt konnte eine Steigerung der Erkennungsrate von 78.4% auf 88.1% im Vergleich zu bisherigen Verfahren erreicht werden.

7 Zusammenfassung

Ziel der in der Dissertation entwickelten Verfahren ist die Reduzierung der Anzahl von notwendigen Lernbeispielen bei der visuellen Objekterkennung. Dabei wurden mehrere Verfahren entwickelt, welche das Konzept des Lerntransfers beim maschinellen Lernen umsetzen. Grundgedanke ist hierbei die Ausnutzung von Lerndaten bereits bekannter Objektkategorien. Weiterhin wurden neue Methoden der Ein-Klassen-Klassifikation vorgestellt, welche bei der Defektlokalisierung, Aktionserkennung und Bildkategorisierung zum Einsatz kommen.

Die beschriebenen Verfahren sind notwendig, um ein effizientes, kontinuierliches und inkrementelles Lernen zu ermöglichen. Dieser Bereich wird aufgrund der immer höheren Anforderungen an automatisch bestimmte, semantische Information zunehmend an Bedeutung gewinnen.

Literatur

- [BCW08] Edwin Bonilla, Kian Ming Chai und Chris Williams. Multi-task Gaussian Process Prediction. In *Advances in Neural Information Processing Systems*, Seiten 153–160. MIT Press, 2008.
- [Bie87] I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychol Rev*, 94(2):115–147, Apr 1987.
- [DDS⁺09] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li und L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, Seiten 248 – 255, 2009.
- [DWSP11] P. Dollar, C. Wojek, B. Schiele und P. Perona. Pedestrian Detection: An Evaluation of the State of the Art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6(99):1030–1037, 2011.
- [FRD10] Björn Fröhlich, Erik Rodner und Joachim Denzler. A Fast Approach for Pixelwise Labeling of Facade Images. In *Proceedings of the 2010 International Conference on Pattern Recognition (ICPR'10)*, Jgg. 7, Seiten 3029–3032, 2010.
- [KRD10] Michael Kemmler, Erik Rodner und Joachim Denzler. One-Class Classification with Gaussian Processes. In *Proceedings of the Asian Conference on Computer Vision*, Jgg. 2, Seiten 489–500, 2010.
- [RD10] Erik Rodner und Joachim Denzler. One-Shot Learning of Object Categories using Dependent Gaussian Processes. In *Proceedings of the 2010 Annual Symposium of the German Association for Pattern Recognition (DAGM'10)*, Seiten 232–241, 2010.
- [RD11] Erik Rodner und Joachim Denzler. Learning with Few Examples for Binary and Multiclass Classification Using Regularization of Randomized Trees. *Pattern Recognition Letters*, 32(2):244–251, 2011.
- [RW05] Carl Edward Rasmussen und Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- [TD04] David M. J. Tax und Robert P. W. Duin. Support Vector Data Description. *Machine Learning*, 54(1):45–66, 2004.

Erik Rodner studierte Informatik mit Nebenfach Mathematik an der Friedrich-Schiller Universität Jena und erwarb sein Diplom im Jahr 2007 mit Auszeichnung. Im Rahmen seiner Promotion am Lehrstuhl für digitale Bildverarbeitung bei Prof. Joachim Denzler, studierte er die Problemstellung des Lernens mit wenigen Beispielen, welche im vorliegenden Paper kurz zusammengefasst wurde. Seine Dissertation wurde im Jahr 2011 einheitlich mit summa cum laude bewertet und erhielt den Promotionspreis der Universität Jena. Aktuell ist er als PostDoc beschäftigt und erforscht Verfahren der visuellen Objekterkennung.

