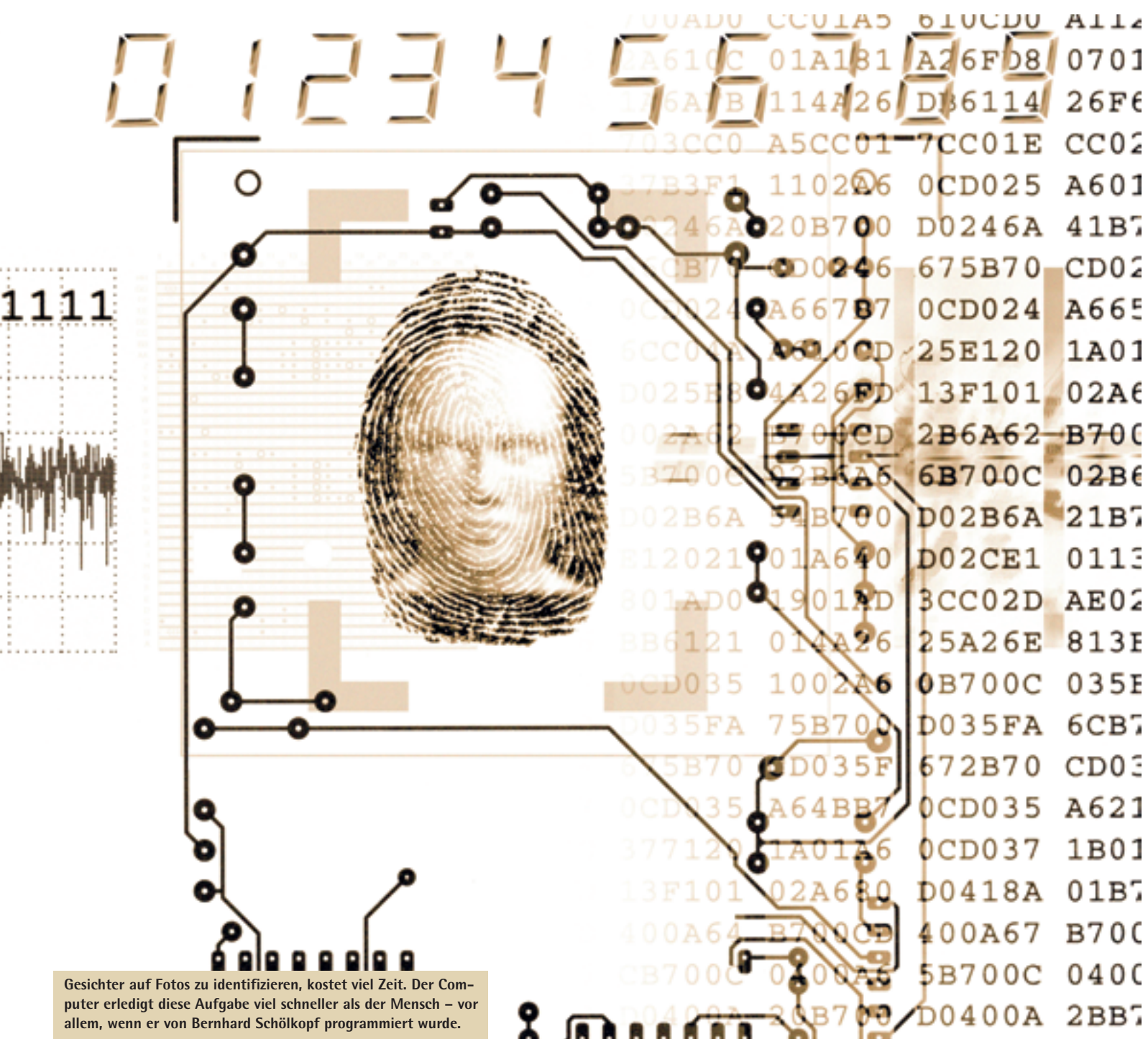


Wo Maschinen in die Schule gehen

Ihre Studienobjekte sind Computerprogramme und Rechenalgorithmen – Maschinen, die komplexe Probleme lösen sollen und deren Stärke in der Verarbeitung großer und unübersichtlicher Datenmengen liegt. Daher befassen sich die Lerntheoretiker um **BERNHARD SCHÖLKOPF**, Direktor am **MAX-PLANCK-INSTITUT FÜR BIOLOGISCHE KYBERNETIK** in Tübingen, nicht mit Schule und Pädagogik, sondern mit der Frage: Wie lernen Maschinen?



Gesichter auf Fotos zu identifizieren, kostet viel Zeit. Der Computer erledigt diese Aufgabe viel schneller als der Mensch – vor allem, wenn er von Bernhard Schölkopf programmiert wurde.

Mit Hebeln und Kolben, mit Fließbändern und Schmieröl haben die Maschinen, mit denen die Wissenschaftler um Bernhard Schölkopf arbeiten, nichts zu tun. Ebenso wenig geht es bei ihren Forschungen um thermodynamische Wirkungsgrade, an die manch einer vielleicht noch bei dem Wort Maschine denkt. Maschinen sind in diesem Fall Computerprogramme oder genauer: Rechenverfahren, so genannte Algorithmen, die von einem Roboter oder Programm ausgeführt werden – oft viele Male hintereinander. Für diese Maschinen braucht man also keine Techniker oder Handwerker.

Aber das Einsatzgebiet lässt auf den ersten Blick auch nicht unbedingt ein Betätigungsfeld für Mathematiker, Informatiker, Ingenieure und Kognitionsforscher vermuten: Statistische Lerntheorie und empirische Inferenz heißt die Abteilung des Tübinger Max-Planck-Instituts für biologische Kybernetik. Und wenn Bernhard Schölkopf über sein Fachgebiet spricht, ist von Training für die Maschinen die Rede, von Entscheidungen, die diese treffen müssen, und von anderen Begriffen, die eigentlich eher nach Pädagogik klingen.

Dabei lässt sich die Grundaufgabe der Lerntheorie ganz naturwissenschaftlich-nüchtern formulieren: Aus

einer Reihe von Beobachtungen – das können beliebige Messwerte oder Daten sein – soll eine Gesetzmäßigkeit gefolgert werden. Diese muss nicht nur erklären, wie die schon vorliegenden Beobachtungen zusammenhängen, sondern auch zukünftige Beobachtungen zuverlässig vorhersagen. Das ist wie in einer Denksportaufgabe: Welche Zahl muss man hinzufügen, um die Folge 3 – 4 – 6 – 10 – 18 logisch fortzusetzen? Etwas Nachdenken und Kopfrechnen führt auf die Lösung: Von jeder Zahl eins abziehen und dann mit zwei multiplizieren; so steht am Ende die 34.

Die Lerntheorie lässt sich auch mit der Wettervorhersage vergleichen: An Stationen von Sylt bis zur Zugspitze werden Daten wie Temperatur, Luftdruck und Niederschlag gesammelt. Daraus berechnen die Meteorologen die jeweils aktuelle Wetterlage und versuchen (mit wechselndem Erfolg), die zukünftige Entwicklung des Wetters vorauszusagen. Ein Algorithmus in einem Computerprogramm geht, mit geeigneten Programmierbefehlen, geradezu menschlich vor. Wie der Mensch lernt auch die Maschine aus Beispielen, wenngleich die Aufgabenstellungen natürlich ganz andere sind.

Traditionell stellen sich in den Naturwissenschaften Induktionsproble-

me: Dabei muss vom besonderen, experimentell zugänglichen Modellfall auf die allgemeine Gesetzmäßigkeit geschlossen werden, auf der dieser basiert. Die Probleme lassen sich lösen, indem ein System so detailliert studiert wird, dass die darin ablaufenden wichtigen Vorgänge aufgedeckt werden. So lässt sich ein Modell aufstellen und überprüfen. Das leisten auch die gelehrigsten Maschinen nicht – Zusammenhänge können sie nicht herleiten. Maschinen eignen sich aber oft dazu, Systeme und Phänomene zu beschreiben, die viel zu komplex für ein mechanistisches Modell sind.

Die Aufgabe für eine Maschine kann etwa darin bestehen, eine medizinische Diagnose zu stellen. Die Maschine schließt also aus einer Reihe von Symptomen auf das Vorhandensein einer bestimmten Krankheit. Außerdem setzen Bioinformatiker Maschinen ein, um das menschliche Proteom zu analysieren, um die Struktur und Wirkungsweise der Proteine zu bestimmen, die der Körper anhand der Blaupause DNA herstellt. Oder aber einer Maschine wird beigebracht, eine Stimme zu identifizieren oder ein Muster in einer Anzahl von Bildpunkten zu erkennen.

Die Gesetzmäßigkeit, die eine Maschine lernen soll, ist an eine Reihe wichtiger Voraussetzungen ge-

Maschinen lernen: Aus den Trainingsdaten zur Klassifikation von Gesichtern (links) wird eine Hypothese abgeleitet, die nur von speziellen Stützpunkten abhängt, den Support-Vektoren (Mitte), welche die Klassifikationsregel beschreiben. Neue synthetisierte Stützpunkte (rechts) enthalten die selbe Regel, allerdings kompakter als zuvor, und erlauben somit schnellere Entscheidungen.

FOTOS: SCIENCE PHOTO LIBRARY/ MPI FÜR BIOLOGISCHE KYBERNETIK

knüpft: Sie muss – zumindest näherungsweise – die vorgegebenen Beispiele erklären, das heißt, das empirische Risiko muss gering sein. Außerdem sollte die Erklärung einfach sein und sich auch auf zukünftige Beobachtungen anwenden lassen. Die Wissenschaftler bezeichnen das als Generalisierungsfähigkeit. Die Abbildung unten zeigt, dass ein niedriges empirisches Risiko nicht gleichbedeutend ist mit hoher Generalisierungsfähigkeit: Punkte symbolisieren die Beispiele, Linien die gesuchte Gesetzmäßigkeit. Die Kurve enthält alle Beispiele, die Gerade führt knapp an den Punkten vorbei. Trotzdem ist man intuitiv geneigt, der Geraden mehr Vertrauen zu schenken. Taucht in der Abbildung ein neuer Punkt (eine neue Beobachtung, ein neuer Messwert) auf, so muss dieser immer noch möglichst nahe an der gesuchten Funktion liegen.

Wo der neue Messpunkt erscheint, darüber gibt es natürlich keine sicheren Kenntnisse, sondern nur eine vorgegebene Wahrscheinlichkeitsverteilung. Wieder hilft der Vergleich mit einer Wettervorhersage: Die Wahrscheinlichkeit, in Köln im August Temperaturen unterhalb des Gefrierpunkts oder oberhalb von 100 Grad Celsius zu messen, ist praktisch null. Nur etwas größer ist die Wahrscheinlichkeit für Temperaturen von 5 oder von 75 Grad Celsius. Wahrscheinlich wird die Temperatur zwischen 10 und 40 Grad Celsius liegen. Der Ort des Beispielpunkts soll ja gerade durch die ge-

suchte Gesetzmäßigkeit möglichst genau vorhergesagt werden.

Die Maschine, die anhand von Trainingsbeispielen die zunächst unbekannte Gesetzmäßigkeit gefolgert hat, wird also mithilfe von Testbeispielen auf ihre Fähigkeit überprüft, zukünftige Beobachtungen vorherzusagen. Diese Generalisierungsfähigkeit ist das wesentliche Qualitätskriterium – eben genau wie bei der Wettervorhersage. Ob das Zentrum des Tiefs, das für den letzten Sturm verantwortlich war, über Island oder dem Skagerrak lag, ist allenfalls für Meteorologen interessant. Der Landwirt oder Bergsteiger möchte wissen, wie es sich auf das Wetter in den kommenden Tagen auswirkt. Bernhard Schölkopf sagt dazu: „Als Naturwissenschaftler erwartet man von einem Modell, dass es Einsicht in die zu Grunde liegenden Phänomene vermittelt. Bei der Analyse komplexer hochdimensionaler Probleme ist das nicht immer möglich – und so wird die Qualität eines Modells in der Lerntheorie vor allem nach seiner Generalisierungsfähigkeit beurteilt, also nach einer Größe, die *a priori* nichts mit Einsicht zu tun hat.“

COMPUTER ZUR GESICHTSKONTROLLE

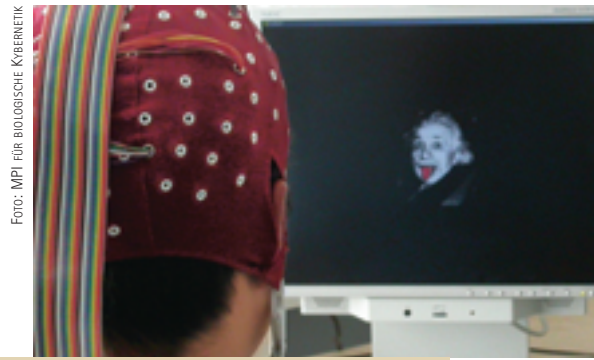
Allgemein muss eine Maschine zukünftige Beobachtungen klassifizieren. Das heißt, abhängig davon, wie die vorhergesagte Beobachtung ausfällt, muss sie die Werte in Klassen einteilen. Bei der binären Mustererkennung beschränkt sich die Vorhersage auf eine Entscheidungsfunktion, die nur die zwei Werte „Ja“ oder „Nein“ annehmen kann. Eine Maschine, der man beigebracht hat, ein bestimmtes Muster zu erkennen, muss für unbekannte Testbeispiele eine korrekte Klassifikation liefern. Sie muss also entscheiden, ob das Testbeispiel das fragliche Muster – etwa ein Dreieck, einen Gegenstand oder einen bestimmten Buchstaben – darstellt oder nicht.

Diese Aufgabenstellung hat ganz praktische Anwendungsmöglichkeiten, zum Beispiel die automatische

Schrifterkennung. Oder aber eine andere, in der die Wissenschaftler um Schölkopf kürzlich einen wichtigen Fortschritt erzielt haben: die computerunterstützte Gesichtsdetektion. Überwachungskameras sind nämlich ein wichtiges Hilfsmittel bei der Terrorabwehr; ihre Aufnahmen auszuwerten kostet aber viel Zeit und Konzentration. Schölkopf und seine Forscherkollegen liefern jetzt mit einem neuen Verfahren Ergebnisse, die wesentlich dabei mithelfen, diese Arbeit einem Computer zu übertragen.

In einem Aufsatz für die *PROCEEDINGS* der britischen Royal Society beschreiben die Max-Planck-Forscher eine Methode, mit deren Hilfe Computer erheblich schneller als bisher Gesichter auf Fotos oder Aufnahmen von Internetkameras zu finden vermögen. Sie beruht auf einem Verfahren, das in der statistischen Datenauswertung häufig angewandt wird, dem so genannten Support-Vektor-Verfahren. Das wird auch in der Medizin eingesetzt, um Gewebeproben eines Patienten zu untersuchen. Anhand der Genaktivität soll dann verlässlich entschieden werden, ob der Patient an einer bestimmten Krankheit wie Leukämie oder einer anderen Krebsart leidet. Die Trainingsbeispiele sind in diesem Fall die Gen-Expressionsprofile von Patienten mit bekannter Diagnose. Ziel ist die korrekte Diagnose für das Expressionsprofil (die Genaktivität) eines neuen Patienten, dessen Diagnose noch nicht feststeht.

Entwicklungsbiologen benutzen das Support-Vektor-Verfahren, um Einzelheiten über die Genaktivität während der Entwicklung von *Drosophila*-Embryonen herauszufinden. In der Terrorabwehr versucht man damit festzustellen, ob und wo eine große, komplexe Kameraaufnahme Gesichter enthält. Um diese Frage zu beantworten, teilt man das Bild in einzelne Abschnitte auf. Für diese Abschnitte muss jeweils entschieden werden, ob die untersuchten Bildpunkte die Abbildung eines Gesichts enthalten. Damit ist freilich nur ein Teil der Aufgabe bewältigt, Bilder automatisch



Schnittstelle Mensch – Maschine: Ein Proband stellt sich vor, seine Zunge zu bewegen. Ein lernendes Computerprogramm erkennt die für dieses innere Bild charakteristischen EEG-Signale.

auszuwerten. Mit der Vorgehensweise der Mathematiker werden Gesichter nur gefunden, nicht identifiziert. Dennoch bedeutet das eine erhebliche Erleichterung und Beschleunigung bei der Bildauswertung.

Die Support-Vektoren sind aussagekräftige Trainingsbeispiele, die das Computerprogramm nach einer bestimmten Vorschrift berechnet. Sie legen gewissermaßen eine Trennlinie zwischen Gesichtern und Nicht-Gesichtern fest. Mit ihrer Hilfe wird einem Computerprogramm, das Bilder auswertet – einer Support-Vektor-Maschine – in einem mathematischen Verfahren beigebracht, ob es sich bei einem Bildausschnitt tatsächlich um ein Gesicht handelt oder beispielsweise nur um einen hellen Fleck im gemusterten Hintergrund. Gesichter können ganz verschieden groß sein: detaillierte Porträts oder winzige Teile eines Körpers. Außerdem muss ein Gesicht in einer Aufnahme gefunden werden – egal, ob die abgebildete Person einen Schnurrbart oder eine Brille trägt, ob sie nach rechts, nach links oder in die Kamera schaut. Mittels eines mathematischen Tricks lässt sich jeweils die Ähnlichkeit zwischen zwei Bildausschnitten berechnen. An zwei Stellen konnten Schölkopf und seine

Kollegen das Support-Vektor-Verfahren verbessern. Zum einen haben die Wissenschaftler einen Weg gefunden, dabei mit erheblich weniger Support-Vektoren auszukommen.

Die Entscheidung, ob ein Bildausschnitt ein Gesicht enthält, wird vereinfacht. Dennoch muss sie natürlich genau bleiben, das heißt, es darf weder ein Gesicht übersehen noch ein Fehlalarm gegeben werden. Schon die Verwendung eines einzelnen reduzierten Satzes von Support-Vektoren bedingt einen wesentlich geringeren Rechenaufwand. Im Vergleich zur Auswertung mit sämtlichen Support-Vektoren ist so eine 30fach schnellere Bildauswertung möglich.

Zum anderen werden die Support-Vektoren nicht für alle Teile des Bildes gleich aufwändig berechnet. Stattdessen wenden die Wissenschaftler eine „Auswertungs-Kaskade“ an: Sie berechnen dabei mehrere, jeweils reduzierte Sätze von Support-Vektoren. Um einen unspezifischen Hintergrund – etwa ein Wandstück, die Kleidung der abgebildeten Personen oder ein großes Fenster, durch das der Himmel scheint – zu klassifizieren, reichen ein oder zwei Support-Vektoren aus. Für gesichtsähnliche Bildausschnitte braucht es eine komplexere Auswertung, die einigen Dutzend Support-Vektoren entspricht. Zudem werden in jedem Auswertungsschritt die Support-Vektoren des vorigen Schritts in einer Art Recycling wiederverwertet.

Mit der Auswertungs-Kaskade lässt sich der Rechenaufwand also besser konzentrieren. Gesichter werden nur in den Teilen eines Bildes gesucht, in

denen sie – laut Anfangsauswertungen – vorkommen können. Die Kaskade, also die Verwendung einer Folge mehrerer reduzierter Sätze von Support-Vektoren, gestattet im Vergleich zum ursprünglichen Verfahren eine 900-fach schnellere Bildauswertung. Das reicht den Wissenschaftlern aber nicht. Sie streben danach, die Geschwindigkeit weiter zu steigern. Das wollen sie beispielsweise dadurch erreichen, dass sie die Auswahl eines reduzierten Satzes von Support-Vektoren optimieren und diese schneller mit den Ausschnitten der zu untersuchenden Aufnahme vergleichen. Aber nicht nur mit optimierten Berechnungsverfahren, sondern auch durch die Entwicklung schnellerer Computerprozessoren erhoffen sich die Forscher eine noch raschere Bildauswertung – die im Übrigen längst nicht die einzige alltagsnahe Anwendung der statistischen Lerntheorie darstellt.

Eine andere könnte das Leben schwer kranker Menschen erleichtern, die nach einem Schlaganfall oder aufgrund der seltenen neurologischen Krankheit *Amyotrophe Lateralsklerose* (ALS) vollständig gelähmt sind. Solche Patienten können weder sprechen noch andere Muskeln, zum Beispiel die Blickmotorik, einsetzen, um ein Hilfsmittel zu steuern. Aber ihr Gehör funktioniert im Allgemeinen noch gut. Ein Ansatz besteht darin, die Gehirnströme des Patienten aufzuzeichnen, während dieser seine Aufmerksamkeit auf Gehörtes konzentriert. An der Gehirn-Computer-Schnittstelle werden die Gehirnströme dann mit einem Support-Vektor-Verfahren umgewandelt – und eine Kommunikationshilfe gibt schließlich das Wort „Ja“ oder „Nein“ von sich.

STEFANIE HENSE



Anhand vorgegebener Beispiele (blaue Punkte) soll eine unbekannte Gesetzmäßigkeit gefolgert werden. Welche der beiden abgebildeten Hypothesen (Kurve oder Gerade) scheint plausibler? Wenn ein neues Beispiel (grüner Kreis) auftaucht, sollte auch dieses möglichst nahe an der neuen Funktion liegen.

Der Weg zum guten Foto: Links die vierfache Vergrößerung des Originals ohne Detailinformationen. Das mittlere Bild enthält plausible Einzelheiten wie Blätter oder Äste, die aus der statistischen Analyse einer großen Bilddatenbank gewonnen wurden. Rechts hat der Computer diese Details hinzugefügt und das Foto rekonstruiert.

