## Das Geheimnis des Sehens

Das Geheimnis des Sehens liegt in der parallelen Verarbeitung der vielen Sinneseindrücke, die die unterschiedlichen Stellen eines Bildes hervorrufen. Je nach den Notwendigkeiten der verschiedenen Lebewesen hat die Natur viele Formen von Augen entwickelt. So können die Facettenaugen von Insekten zwar nicht so feine Strukturen erkennen wie die Linsenaugen des Menschen, haben dafür aber ein erheblich höheres zeitliches Auflösungsvermögen. Während der Mensch nur bis zu 20 Einzelbilder pro Sekunde unterscheiden kann, sind es bei schnell fliegenden Insekten wie Libellen oder Bienen bis zu 300.

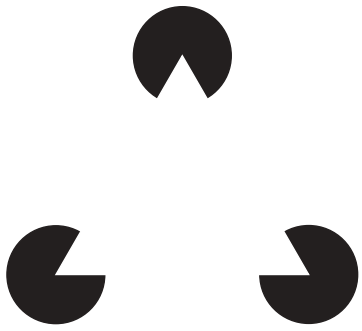
Diese Einzelbilder bestehen beim Libellenauge jedoch aus lediglich etwa 30.000 Bildpunkten. Im menschlichen Auge dagegen wird durch die Linse ein Bild auf die Netzhaut projiziert, wo es von vielen Millionen Sinneszellen verarbeitet wird. Über sechs Millionen Zapfen regeln die Farbwahrnehmung und weit über 100 Millionen Stäbchen sind für die Unterscheidung von hell und dunkel verantwortlich. Tatsächlich handelt es sich also bei der visuellen Wahrnehmung um viele einzelne, gleichzeitige Sinneseindrücke, wobei jede einzelne Sinneszelle nur einen kleinen Bereich des gesamten Bildes wahrnimmt. Dabei sind die Sinneszellen nicht gleichmäßig verteilt. Im Zentrum unseres Sichtfeldes sind sie am dichtesten gepackt (160.000 Zapfen pro Quadratmillimeter), daher können wir dort am schärfsten sehen. Die Randbereiche werden aber stets mitkontrolliert. Wenn dort etwas passiert, kann sich der Blick diesem Bereich zuwenden.

Im Vergleich zur Schaltzeit elektronischer Bauelemente arbeiten unsere Nerven extrem langsam. Trotzdem können wir Bilder viel besser auswerten als künstliche visuelle Systeme. Das gelingt durch die gleichzeitige Verarbeitung der Sinneseindrücke in den einzelnen Nerven. Denn es genügt ja nicht, dass jede Nervenzelle für sich allein arbeitet, entscheidend ist vielmehr die Zusammenfassung der vielen einzelnen Signale zu

einer Wahrnehmung: Alle Sinneszellen, die Teile des Balles registriert haben, müssen zusammen zu dem Resultat kommen, dass sie den Ball sehen. Da darf es nicht stören, wenn sich auf dem Ball ein Schmutzfleck befindet oder Teile verdeckt sind. Die Gesamtwahrnehmung soll den Ball erkennen, selbst wenn einige der Sinneszellen eigentlich nicht glauben können, einen Ball zu sehen.

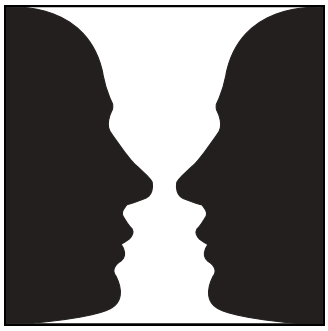
### Demokratie der Sinne

Wahrnehmung erfolgt daher als eine Art Mehrheitsentscheidung: Wir erkennen einen Ball (oder glauben, ihn zu erkennen), wenn hinreichend viele Stimmen dafür sprechen. Diese Stimmen werden durch unsere Erfahrungen und unser Wissen über die Welt beeinflusst: Wir nehmen einen Stuhl wahr, auch wenn wir nur eine Lehne an einem Platz sehen, wo üblicherweise Stühle stehen können. Diese eigentlich voreilige Wahrnehmung lässt sich natürlich täuschen. Zum Beispiel neigen wir dazu, unterbrochene Linien zu vervollständigen und als ganzes wahrzunehmen. Unsere Wahrnehmung rekonstruiert die gesamte Linie. Das ist meistens sinnvoll, weil die Unterbrechungen in der Regel durch davor befindliche Gegenstände verursacht werden. Es kann aber auch dazu führen, dass wir Dinge zu sehen glauben, die nicht vorhanden sind, zum Beispiel ein helles Dreieck vor dunklerem Hintergrund.

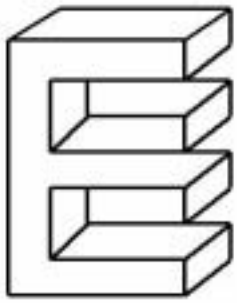


Im Alltag sind diese Fähigkeiten zur Rekonstruktion notwendig, und die Ergebnisse sind meistens auch korrekt, die optischen Täuschungen sind die Ausnahme. Sie verweisen daher nicht auf eine Unzulänglichkeit unserer Wahrnehmung, sondern resultieren aus einem Mechanismus, der normalerweise sehr hilfreich ist.

Insgesamt sind Sehen und Wahrnehmen sehr komplizierte Prozesse, bei denen sich viele einzelne Stimmen eine Meinung bilden müssen. Nur ein Teil dieser Stimmen kommt direkt aus den Sinneszellen im Auge. Deren Meinungen können sogar überstimmt werden. Eigentlich ist es auch keine direkte Abstimmung, sondern eher ein Aushandeln und Weiterleiten von Meinungen. Manchmal kommt keine einheitliche Meinungsbildung zustande, wie bei dem Bild, das eine weiße Vase oder zwei einander zugewandte Gesichter zeigt.



Noch beunruhigender ist das im Fall des Balkens mit den drei oder vier Querbalken. Man kann den Kampf zwischen den beiden Meinungen beim Betrachten des Bildes körperlich spüren. Wenn man die eine Seite des Bildes verdeckt, fühlt man sich sofort wohler.



## Wie Computer lesen lernen

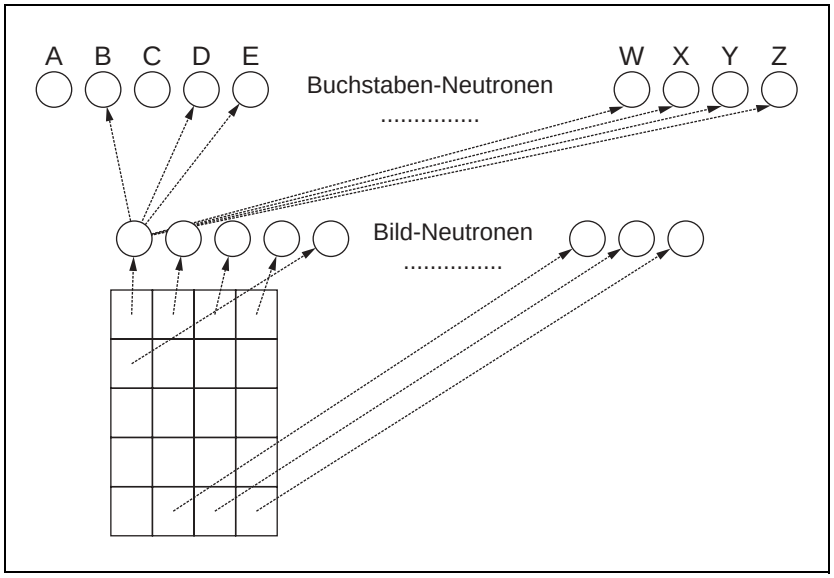
Heutige Bildverarbeitungssysteme arbeiten häufig mit so genannten Künstlichen Neuronalen Netzen (KNN), die die »Abstimmung« der Nervenzellen im Gehirn im Computer simulieren. Ein einfaches Beispiel ist die Erkennung handgeschriebener Großbuchstaben: Die Buchstaben stehen dabei an einer ganz bestimmten Stelle (auf elektronisch zu verarbeitenden Formularen sind diese Bereiche oft auch schön gleichmäßig vorgegeben, manchmal sogar mit Hilfslinien zum genaueren Schreiben). Eine Kamera erzeugt nun ein schwarz-weißes Bild von einem Buchstaben aus genau einem solchen Bereich, und die Software soll den richtigen Buchstaben identifizieren.

Das Bild besteht aus einzelnen Bildfeldern, so genannten Pixeln (abgekürzt aus dem englischen Begriff »picture element«), die je nach Art des Buchstabens heller oder dunkler sind. Die gemessenen Helligkeitswerte können durch Zahlen zwischen 0 und 255 bezeichnet werden, wobei »255« für schwarz steht, »0« für weiß. »118« wäre dann ein mittleres Grau, und »201« steht für ein schon ziemlich dunkles Grau, das heißt, hier hat der Stift ziemlich viel Farbe hinterlassen (in der Bildverarbeitung ist die Kodierung üblicherweise umgekehrt, aber das ist hier völlig nebensächlich).

Jedes Pixel wird in dem neuronalen Netz, das den Buchstaben erkennen soll, durch ein »Bild-Neuron« repräsentiert. Das Erkennen erfolgt durch Abstimmung der Neuronen, wobei die Bild-Neuronen unterschiedlich viele Stimmen haben können. Die Anzahl der Stimmen hängt von der Graufärbung des jeweiligen Pixels ab. Ein Bild-Neuron über einem schwarzen Feld besitzt 255 Stimmen, über dem dunkleren Grau gibt es 201, und über dem Weiß gar keine. Das ist nicht die einzige Verzerrung des Wahlrechts: Je nach Wichtigkeit der betreffenden Bildfelder für die Unterscheidung der Buchstaben werden die Stimmen zusätzlich um unterschiedliche Faktoren vermehrt, einige zum Beispiel verdoppelt, andere vielleicht verzehnfacht. Bei manchen Anwendungen kann einem Neuron auch das Wahlrecht ganz entzogen werden, wenn es eine Mindestzahl von Stimmen unterschreitet. Mit den verfügbaren Stimmen führen die Bild-Neuronen nun eine Abstimmung über die Interpretation des Bildes durch.

Bei der Wahl hat jedes Bild-Neuron einen festen Schlüssel, nach dem es seine Stimmen auf die Kandidaten, die Buchstaben-Neuronen, verteilen muss. So wird das Neuron des oberen linken Eckpunktes seine Stim-

men vorrangig an die Buchstaben »B«, »D« »E«, »F«, »H«, K«, »L«, »M«, »N«, »P«, »R«, »T«, »U«, »V« , »W«, »X«, »Y« und »Z« vergeben, weil bei ihnen dieses Feld in der Regel beschrieben ist. Bei etwas eckiger Schreibweise könnten allerdings auch »A«, »C«, »G«, »O«, »Q« oder »S« dort eine Graufärbung verursachen. Also bekommen diese Buchstaben ebenfalls einige Stimmen von dem Bildpunkt, wenn auch nicht ganz so viele. Das rechte untere Eckfeld müsste dementsprechend seine Stimmen vorrangig an die Buchstaben »A«, »E«, »H«, »K«, »L«, M«, »N«, »R«, »X«, und »Z« verteilen. Das sind deutlich weniger als beim linken oberen Eckfeld – was ein Grund dafür sein könnte, die Stimmen des linken oberen Eckfeldes vor der Abstimmung zu vervielfachen.




Angenommen, der Stift des Schreibers habe das linke obere Eckfeld zur Hälfte beschrieben, und die Kamera habe eine mittlere Graufärbung registriert. Das Neuron habe deshalb zunächst 150 Stimmen zur Verfügung. Weil es ein sehr wichtiges Neuron ist und seine Stimmen auf viele Buchstaben verteilen muss, wird seine Stimmenzahl um den Wichtigkeitsfaktor vier erhöht. Das Neuron hat jetzt also 600 Stimmen. Der Schlüssel zur Aufteilung sei für das linke obere Eckfeld so festgelegt, dass die Neuronen der oben genannten Buchstaben »B«, »D« »E«, »F«, »H«, K«, »L«, »M«, »N«, »P«, »R«, »T«, »U«, »V« , »W«, »X«, »Y« und »Z« jeweils 5 % der Stimmen erhalten, die Neuronen der Buchstaben

»C«, »G«, »O«, »S« jeweils 2 % und die der Buchstaben »A« und »Q« jeweils 1 %. Die übrigen Buchstaben erhalten nichts. In unserem Fall erhielte der Buchstabe »L« also 150 * 4 * 0,05 = 30 Stimmen. Entsprechend der jeweils vorliegenden Graufärbung, des Wichtigkeitsfaktors und des Verteilungsschlüssels erhält der Buchstabe »L« auch Stimmen von den anderen Bild-Neuronen, alle diese Stimmen werden summiert. Dann werden die Wahlergebnisse miteinander verglichen, und der Wahlsieger ist der erkannte Buchstabe.

Das gleiche Verfahren wird für die anderen Buchstaben auf dem Formular wiederholt. Jedes Mal ist dabei die Graufärbung anders verteilt, und jedes Mal ergibt die Auszählung eine andere Stimmenverteilung. Wenn die Wichtigkeitsfaktoren und die Verteilungsschlüssel der einzelnen Bild-Neuronen gut gewählt sind (sie sind immer die gleichen), erhält der richtige Buchstabe die meisten Stimmen. Voraussetzung ist natürlich, dass der Buchstabe hinreichend deutlich geschrieben ist. Ein undeutliches »P« mit tief ansetzendem Bogen kann auch ein Mensch kaum von einem »D« unterscheiden.

Damit das alles korrekt funktioniert, müssen die Wichtigkeitsfaktoren und die Verteilungsschlüssel richtig gewählt werden. Mathematisch kann man sie zu einer Zahl zusammenfassen, dem Gewicht der Verbindung zwischen zwei Neuronen (das Gewicht ist einfach das Produkt der beiden). Jetzt können wir das auch zeichnen: Die Neuronen sind Kreise, und von jedem Eingangsneuron wird ein Pfeil zu jedem Ausgangsneuron gezeichnet, an den man das Gewicht schreibt. Pfeile mit dem Gewicht Null lässt man natürlich weg.

Damit erhält man ein einfaches Künstliches Neuronales Netz (KNN). Anstatt von Stimmen wird dann von »Aktivierungen« der Neuronen gesprochen, und anstelle von Abstimmungen betrachtet man die Ausbreitung von Aktivierungen durch das Netz:

Am Anfang sind nur die Bild-Neuronen (die »Eingangs-Neuronen«) aktiviert. In unserem Fall hängt ihre Aktivierung von der Graufärbung ab, in anderen Anwendungen kann das etwas ganz anderes sein. Dann schicken sie Impulse längs der Pfeile an die Buchstaben-Neuronen (»Ausgangs-Neuronen«). Die Stärke der Impulse ist abhängig von der Aktivierung und den jeweiligen Gewichten (zum Beispiel Impuls = Aktivierung · Gewicht). An den Buchstaben-Neuronen werden alle ankommenden Impulse zu einer Aktivierung zusammengefasst (zum Beispiel als Summe der Impulse).

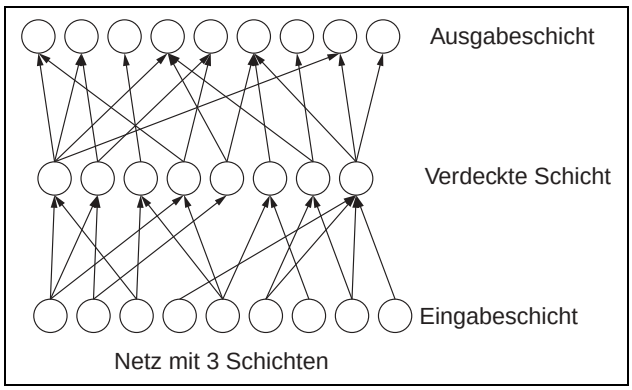
### Analogie zum Gehirn

Künstliche Neuronale Netze tragen ihren Namen, weil sie an die Verschaltung von Neuronen in organischen Gehirnen erinnern. Auch die Bezeichnung »Aktivierung« stammt von dort: Man kann sich die Arbeitsweise so vorstellen, dass Aktivierungen in den Neuronen verstärkt werden (das waren ursprünglich bei uns die Wichtigkeitsfaktoren) und dann entlang der Verbindungen in unterschiedlicher Stärke (das waren die Verteilungsschlüssel) weitergeleitet werden. Mathematisch kann man das wie schon beschrieben allein durch die Gewichte an den Verbindungen modellieren. Dabei kann es auch negative Werte geben: In der Abstimmung wären das »Nein«-Stimmen, die bei der Auszählung gegen die »Ja«-Stimmen gesetzt werden. Bei der Ausbreitung von Aktivierungen bewirkt das eine Hemmung des empfangenden Neurons.

Die Ausgangsneuronen eines Netzes können nun wiederum als Eingangsneuronen eines weiteren Netzes fungieren. Im Beispiel des Formulars könnten die Buchstaben-Neuronen die Eingangs-Neuronen eines Netzes sein, das Wörter erkennt. In einer einfachen Form hätten wir für jedes Buchstabenfeld des Formulars ein Buchstabenerkennungsnetz. Die Buchstaben-Neuronen dieser Netze leiten ihre Aktivierungen an die Wort-Neuronen weiter, wenn der Buchstabe in diesem Wort vorkommt.

Wir haben jetzt ein KNN mit einer Eingabe-Schicht (Bild-Neuronen), einer inneren Schicht (Buchstaben-Neuronen) und einer Ausgabe-Schicht (Wort-Neuronen). Die innere Schicht heißt auch verdeckte Schicht, weil ihre Neuronen von außen nicht beobachtbar sind.

KNN haben nun eine Besonderheit: Bis zu einem gewissen Grad können sogar falsch erkannte Buchstaben ausgeglichen werden. Hätte die Buchstaben-Erkennung zum Beispiel »Horwart« gelesen, würde das




Wort »Torwart« immer noch die Stimmen von 6 Buchstaben-Neuronen erhalten. Diese Fehlertoleranz ist eine ganz wichtige Eigenschaft auch der natürlichen neuronalen Netze. Am Beispiel der Buchstabenerkennung wird sie noch deutlicher: Ein zusätzlicher Punkt, etwa ein Staubkorn, auf einem Bildfeld muss nicht sofort zu einer Fehlleistung führen. Fehler treten erst dann auf, wenn das Abstimmungsergebnis eines falschen Buchstabens besser wird. Auch beim Ausfall einzelner Verbindungen oder sogar Neuronen kann das Ganze noch recht gut funktionieren.

### Training ist wichtig

Voraussetzung dafür ist, dass sich die Abstimmungsergebnisse, die Aktivierungen, bei den »typischen« Erscheinungsformen der Buchstaben gut voneinander abheben. Das kann dadurch erreicht werden, dass die Verbindungen zwischen den einzelnen Neuronen günstig gewichtet sind. Die geeigneten Werte lassen sich allerdings kaum theoretisch ermitteln. Stattdessen können geeignete Gewichte anhand von Beispielen gelernt (trainiert) werden. Dafür gibt es viele Verfahren.

Wie beim Menschen ist das Trainieren eines KNN ein Lernvorgang, bei dem zu erbringende Leistungen wiederholt geübt und durch einen Beobachter (den Trainer) bewertet werden. Im Falle unseres Buchstabenerkenners würde man das Netz also handgeschriebene Buchstaben untersuchen lassen und nachsehen, ob der Buchstabe richtig erkannt wurde. Nach jedem Durchlauf wird das Netz so verändert, dass die Aufgabe in Zukunft besser gelöst wird. Dazu müssen die Vorgaben für die Stimmenverteilung bei der Wahl verändert werden, das heißt, die Gewichte an den Verbindungen werden variiert. Eine einfache Form ist

das der Natur nachempfundene Hebbsche Lernen (eine um 1950 von D. O. Hebb vorgeschlagene Regel): Wird ein Buchstabe richtig erkannt, werden alle daran beteiligten Verbindungen verstärkt, das heißt, ihre Gewichte werden etwas erhöht. Umgekehrt kann man auch aus Fehlern lernen: Bei den zu hoch bewerteten falschen Buchstaben werden die Verbindungen geschwächt, die zu dieser Bewertung einen hohen Beitrag geleistet haben. Grundsätzlich muss man bei den Veränderungen behutsam vorgehen, weil man sonst schnell wieder zerstören könnte, was vorher bereits gelernt wurde.

Das Trainieren der KNN wird dem Rechner überlassen. Dazu bekommen alle Gewichte zunächst einen Anfangswert. Dann werden die Trainingsbeispiele mit dem Netz durchgerechnet, und je nach dem Ergebnis werden die Gewichte variiert, bis am Ende alle Trainingsbeispiele richtig erkannt werden. Manchmal muss man sich auch damit begnügen, dass wenigstens die meisten Beispiele richtig erkannt werden.

Wie gut und wie schnell das gelingt, hängt natürlich von der gestellten Aufgabe ab. Wir können die Berechnung für optimale Torschüsse einem KNN übertragen. Dabei sollen Werte für die Kick-Richtung und Kick-Stärke in Abhängigkeit von der aktuellen Situation berechnet werden, also aus Werten für den Winkel und die Entfernung zum Tor, der Lage des Balles und der Position des gegnerischen Torwartes. Es gibt Eingangs-Neuronen für alle diese Parameter, und deren Anfangsaktivierungen hängen von den konkreten Werten ab. Dann lässt man das Netz rechnen, das heißt, der Computer rechnet aus, wie sich die Aktivierungen im Netz ausbreiten. Für die Kick-Richtung und die Kick-Stärke gibt es Ausgangs-Neuronen.

Dort liest man zum Schluss die Aktivierungen ab, sie liefern die Werte für den optimalen Kick. Anders als bei der Buchstabenerkennung werden am Ende Zahlenwerte berechnet und nicht nur Auswahlentscheidungen für Buchstaben getroffen.

Auch diese Netze müssen trainiert werden, ehe sie optimale Resultate liefern. Dabei kann es Trainingsfälle geben, in denen der Stürmer zwar optimal geschossen hat, eine plötzliche Windböe den Schuss aber trotzdem am Tor vorbeileitet. Der Trainer würde also für diesen Versuch einen Misserfolg registrieren. Wenn der gleiche Schuss bei einem weiteren Versuch genau in das Tor geht, stehen zwei widersprüchliche Trainingsbeispiele zur Verfügung. Bei gutem Training können KNN aber auch mit solchen Widersprüchen recht gut fertig werden.

Das Training eines KNN kann je nach Schwierigkeit der Aufgabe längere Zeit dauern. Allerdings wird durch längeres Training das Netz nicht immer besser. Es kann passieren, dass es sich auf bestimmte Besonderheiten in der Trainingsmenge spezialisiert. Deshalb müssen mit dem fertig trainierten Netz weitere Testbeispiele durchgerechnet werden. Zeigen diese kein befriedigendes Ergebnis, wird ein neuer Trainingsversuch mit einem neuen Netz gestartet: Die Gewichte erhalten andere Anfangswerte, und das Ganze beginnt von vorn.

### Unter der Ebene der Symbole

Das oben beschriebene Worterkennungssystem ist eigentlich untypisch für Künstliche Neuronale Netze, weil die Neuronen der verdeckten Schicht konkrete Bedeutungen (Buchstaben) besitzen. Bei der Erkennung handgeschriebener Wörter können jedoch noch weitere Gesichtspunkte ausgenutzt werden: Wie ein »A« aussieht, hängt davon ab, ob es auf ein »B« folgt oder ein »V«. Dadurch lassen sich Fehler durch falsch erkannte Buchstaben noch besser vermeiden.

Es ist auch nicht notwendig, dass die inneren Neuronen den Buchstaben entsprechen. Wir brauchen am Eingang die Bilderkennungs-Neuronen und am Ausgang die Worterkennungs-Neuronen. Dazwischen können in der verdeckten Schicht irgendwelche Neuronen sein, über deren Bedeutung wir uns keine Gedanken machen müssen. Was insgesamt im Netz passiert, wird durch die Gewichte bestimmt, die wir im Training festlegen. Indem wir uns davon befreien, dass die inneren Neuronen den Buchstaben entsprechen müssen, können wir beim Training vielleicht Gewichte finden, die viel bessere Ergebnisse liefern, weil sie weitere Zusammenhänge ausnutzen.

Viele dieser Zusammenhänge lassen sich sowieso nur schwer oder gar nicht in Worten (»in Symbolen«) beschreiben. Der prototypische Buchstabe »B« hat zwar bestimmte Eigenschaften wie einen geraden senkrechten Strich und daran ansetzend zwei Halbkreise unterschiedlicher Größe. Ein handgeschriebener Buchstabe »B« kann davon jedoch deutlich abweichen. Er muss nur ungefähr dem Prototypen entsprechen oder im Vergleich mit allen anderen Buchstaben dem »B« am ähnlichsten sehen. Eine solche Ähnlichkeit anhand der Kriterien »senkrechter Strich und daran ansetzend zwei Halbkreise« zu bewerten ist schwierig. Insbesondere müsste die Maschine zuerst wieder in der Lage sein, Striche und

Halbkreise zu identifizieren. Das beschriebene Abstimmungsverfahren der Bild-Neuronen ist da wesentlich effizienter. In Bezug auf die Eigenschaften des prototypischen »B« haben diese Neuronen aber eigentlich keine Benennung. Anders ausgedrückt: Sie tragen keine symbolische Bedeutung.

Man spricht bei KNN deshalb auch von »subsymbolischer« Verarbeitung, von Verarbeitung unterhalb benennbarer symbolischer Bedeutungen. Das Wissen über die Eigenschaften des prototypischen »B« ist zwar in dem Buchstabenerkennungsnetz enthalten, aber es ist nicht an einer festen Stelle kodiert. Vielmehr ist dieses Wissen über das ganze Netz verteilt, nämlich in den Gewichtungen. Gleichzeitig ist dort auch das Wissen über die Eigenschaften der anderen Buchstaben enthalten. Das hat den bereits erwähnten wichtigen Vorteil: Selbst wenn einzelne Verbindungen oder sogar ganze Neuronen ausfallen, reicht das Wissen im verbliebenen Netz meist immer noch zur Lösung der Aufgaben aus, vielleicht bei geringer Erhöhung der Fehlerquote. Wird dagegen in einer Datenbank ein Eintrag gelöscht, ist diese Information komplett verloren. Das ist der entscheidende Vorteil, der durch die verteilte Repräsentation erreicht wird. Hätten wir ein einzelnes spezielles Neuron für den Begriff »Großmutter«, dann würden wir beim Ausfall dieses Neurons nichts mehr von »Großmutter« wissen können. So aber ist der Begriff an vielen Stellen präsent und kann nicht einfach verloren gehen. Man spricht auch von Konnektionismus, um auszudrücken, dass in einem KNN irgendwie alles miteinander verbunden ist.