Hochschule für Technik, Wirtschaft und Kultur Leipzig (FH)
Fachbereich Informatik, Mathematik und Naturwissenschaften

Oberseminar-Arbeit

Das Neocognitron

Vorgelegt von

Matthias Jauernig

Leipzig, Mai 2007

Lehrveranstaltung: Oberseminar Künstliche Intelligenz
Verantwortlicher Professor: Prof. Dr.rer.nat.habil. Siegfried Schönherr

Zusammenfassung

Diese Arbeit beinhaltet die Vorstellung des Neocognitron-Netzwerkes. Das Neocognitron ist eines mehrschichtiges hierarchisches Künstliches Neuronales Netz, welches für Zwecke der Mustererkennung durch den japanischen Forscher Kunihiko Fukushima entwickelt wurde und im Gegensatz zu vielen anderen Systemen die Fähigkeit besitzt, tolerant gegenüber Verschiebungen, Skalierungen und Deformationen zu sein. Es wird zunächst die Architektur des Netzes beschrieben, weiterhin erfolgt die Beschreibung des Prozesses der Mustererkennung, auch unter Berücksichtigung verformter Muster. Nachdem auf die Lernverfahren eingegangen wurde, werden abschließend kurz einige wichtige Modifikationen sowie Anwendungsgebiete des Neocognitrons aufgezeigt.

Inhaltsverzeichnis

1 Einleitung **1**
1.1 Zielsetzung der Arbeit . . . 1

2 Struktur des Netzwerks **2**
2.1 Grundlegender Aufbau . . . 2
2.2 Aufbau und Verbindungsstruktur der Schichten . . . 4
2.2.1 S-Schichten . . . 4
2.2.2 C-Schichten . . . 7
2.3 Arten von Gewichten und Verbindungen . . . 10
2.4 Mathematische Beschreibung . . . 12

3 Prozess der Mustererkennung **14**

4 Erkennung verformter Muster, Verschiebungsunabhängigkeit **15**

5 Lernverfahren **18**
5.1 Unüberwachtes Lernen . . . 18
5.2 Überwachtes Lernen . . . 19

6 Modifikationen des Neocognitrons **21**
6.1 Neocognitron mit *Selective-Attention*-Mechanismus . . . 21
6.2 Doppelte C-Zellenschicht . . . 22
6.3 Rotations-invariantes Neocognitron . . . 23
6.4 Verbessertes Neocognitron . . . 23

7 Anwendungen **24**
7.1 Erkennen von Drahtmodell-Objekten . . . 24
7.2 Objekterkennung . . . 24
7.3 Erkennung handschriftlicher Zeichen . . . 25
7.4 Erkennung koreanischer Schiftzeichen . . . 25
7.5 Extraktion von Symmetrieachsen . . . 25
7.6 Wiederherstellung und Erkennung teilweise verdeckter Muster . . . 26

8 Zusammenfassung und Ausblick **26**

A Glossar **I**

Abbildungsverzeichnis

1 Primitive Muster, mit denen U_{S1} trainiert werden kann (nach [8]) . . . 3
2 Beispiele von zum Training von U_{S2} verwendeter Muster (nach [10]) . . . 3
3 Struktur und grundlegende Verbindungsstrategie des Neocognitrons (nach [1]) . 4
4 Beispielhafte Verbindungen von Zellen der Eingabeschicht U_0 mit U_{S1} . . . 5
5 Beispiel für ein eingegebenes Muster und die Aktivierung von S-Zellen dadurch . 6
6 Verbindungen von V- und S-Zelle sowie Vorgängerschicht U_{Cl-1} . . . 7
7 Aktivierung mehrerer C-Zellen durch eine S-Zelle . . . 8
8 Verarbeitung eines gelernten Musters durch die erste Stufe des Netzes . . . 9
9 Zusammenfassung von S-Ebenen durch eine C-Ebene aufgrund ähnlicher Muster 10

10 Beispiel für die Gewichtsmatrix zwischen S-Zellen und einer C-Zelle 11
11 Übersicht der Verbindungsarten im Neocognitron (nach [8]) 12
12 Beispiel für die Erkennung des Buchstaben `A`durch das Neocognitron (nach [8]) . 15
13 Erkennung von `A`über die 1. und 2. Stufe des Netzes (nach [8]) 16
14 Funktionsweise der C-Zellen zur Erkennung deformierter Muster (nach [16]) . . 17
15 Probleme bei der Deformationserkennung (nach [10]) 17
16 Beispiele für die Erkennung mittels *selective attention* (aus [16]) 22

1 Einleitung

Menschen ist es dank ihres Wahrnehmungssystems leicht möglich, Objekte in ihrer Umwelt und auf Bildern zu erkennen und zu klassifizieren. Dieser Vorgang ist unabhängig von der Größe und Ausrichtung eines Objektes. Selbst teilweise Überdeckungen und Verformungen beeinflussen die Wahrnehmung kaum. Diese Fähigkeiten in ein mathematisches Modell zu überführen und somit auf den Rechner zu übertragen ist Gegenstand jahrzehntelanger Forschung, die bis heute anhält. Vor allem im Bereich der Künstlichen Neuronalen Netze (KNNs) wurden dabei Ergebnisse erzielt, die aus der breitflächigen Mittelmäßigkeit der meisten anderen Lösungen heraustreten konnten und Einzug in praktische Anwendungen fanden. Als ein solches Modell ist das *Cognitron* zu nennen, welches 1975 durch den japanischen Forscher Kunihiko Fukushima entwickelt wurde. Dieses beschreibt als ein mehrschichtiges, selbst-adaptives KNN ein Modell der visuellen Mustererkennung im Gehirn. Allerdings ist es wie viele andere Lösungen auch sensibel gegenüber Verschiebungen, Skalierungen und Deformationen von zu erkennenden Mustern. Aufgrund dessen wurde das Modell von Fukushima weiter angepasst und es entstand ein neues Cognitron, welches 1980 vorgestellt wurde: das *Neocognitron* [1]. Wie auch das Cognitron handelt es sich beim Neocognitron um ein hierarchisches mehrschichtiges Netz, welches dem visuellen System von Säugetieren nachempfunden ist und besonders für den Einsatz bei der Erkennung von Mustern und handschriftlichen Zeichen konzipiert wurde. Nebenbei ging es bei der Entwicklung des Neocognitrons auch darum, durch das damit verbundene Modell des visuellen Sehsystems die Vorgänge im Gehirn besser verstehen zu lernen (siehe [7]) und eine Art künstliches Gehirn zu entwickeln, welches näher am menschlichen Gehirn orientiert ist als bisherige Entwicklungen (siehe [2]). Im Gegensatz zum Cognitron und zu den meisten konventionellen Mustererkennungs-Systemen ([8]) ist das Neocognitron durch seinen speziellen Aufbau bis zu einem bestimmten Grad tolerant gegenüber Verschiebungen, Skalierungen sowie Verformungen und zu einem geringen Teil auch gegenüber Rotationen, was die Haupt-Eigenschaften dieses KNNs sind. R. Hecht-Nielsen stellt es in seinem Buch 1990 (siehe [9]) als das damals einzige wohl erforschte hierarchische KNN und eine der krönenden Entwicklungen im Bereich des Neurocomputings dar und beschreibt es weiterhin als das größte und komplizierteste KNN seiner Zeit.

1.1 Zielsetzung der Arbeit

Diese Arbeit stellt das Neocognitron vor, wie es von Prof. Fukushima entwickelt wurde. Dabei wird besonderer Wert auf die Effekte gelegt, welche sich aus dem speziellen Netzaufbau ergeben und schließlich zu den grundlegenden Eigenschaften des Netzes führen.

In Abschnitt 2 wird die Netzwerk-Architektur beschrieben und neben den Eigenschaften des Netzes erfolgt auch eine mathematische Beschreibung des Ein-/Ausgabe-Verhaltens. Die Abschnitte 3 und 4 gehen auf den Prozess der Mustererkennung ein, bevor Abschnitt 5 die beiden

grundsätzlich verwendeten Lernverfahren skizziert. Die Abschnitte 6 und 7 gehen noch auf einige Modifikationen und Anwendungsgebiete des Neocognitrons ein, um einen abschließenden Rahmen zu geben.

2 Struktur des Netzwerks

Dieser Abschnitt beschäftigt sich mit der Stuktur des Neocognitrons. Unterabschnitt 2.1 klärt zunächst grundlegend über den allgemeinen Aufbau und die Arbeitsweise des Netzes auf, in den folgenden Unterabschnitten wird diese Erst-Beschreibung ausgebaut und mit Details angereichert.

2.1 Grundlegender Aufbau

Ein Neocognitron-Netz besteht aus mehreren miteinander durch selektive Verbindungen gekoppelten *Stufen* (*stages*), durch die Signale in Vorwärtsrichtung geleitet und verarbeitet werden. Bei „Signalen“ handelt es sich dabei um nicht-negative analoge Werte, die als Eingaben und Ausgaben der Zellen dienen ([7]). Charakteristisch für das Neocognitron ist dabei nicht die hohe Anzahl an Zellen, sondern eine hohe Anzahl an Verbindungen mit entsprechenden Gewichtungen zwischen einzelnen Zellen benachbarter Schichten. Die allgemeine Struktur orientiert sich am visuellen System von Katzen und dient als Modell der dort stattfindenden Mustererkennung. So wird bei den Neuronen (den *Zellen*, engl. *cells*) zwischen S- und C-Zellen unterschieden, wie sie auch im primären visuellen Kortex zu finden sind. S-Zellen (*simple cells*) extrahieren Muster-Eigenschaften, während C-Zellen (*complex cells*) leichte Verformungen kompensieren und positionsbedingte Fehler ausgleichen können. Jede Stufe l des Neocognitrons besteht aus einer *Schicht* (*layer*) von S-Zellen (S-Schicht U_{Sl}, *S-layer*) gefolgt von einer Schicht aus C-Zellen (C-Schicht U_{Cl}, *C-layer*). Jede S- und C-Schicht ist dabei weiterhin in mehrere *Ebenen* (*planes*) unterteilt, welche die einzelnen Zellen als rechteckiges Feld enthalten. Eine Ausnahme davon stellt die erste Schicht U_0 dar, welche als *Eingabeschicht* fungiert. Sie besteht aus einem rechteckigen Feld von Rezeptor-Neuronen (*rezeptives Feld*), deren Ausgaben an die Schicht U_{S1} lediglich die Grauwerte des übergebenen Bildes reellwertig oder in binarisierter Form darstellen. Ein grundlegendes Prinzip ist die hierarchische Merkmalsextraktion, wie sie auch von Neuronen des Großhirns durchgeführt wird. Dort reagieren Neuronen des visuellen Kortex auf lokale Eigenschaften wie Linien oder Kanten bestimmter Orientierung. Ausgaben von diesen Neuronen werden eine Ebene höher geleitet, wo Zellen selektiv auf bestimmte Figuren wie Kreise, Rechtecke oder selbst menschliche Gesichter reagieren (siehe [3]). Ebenso werden im Neocognitron verschiedene zu erkennende Merkmale von Zeichen auf mehrere Schichten verteilt. Die Schicht U_{S1} erkennt aus den Eingabe-Signalen primitive Strukturen wie beispielsweise unterschiedliche Linienformen. Dabei wird nur ein sehr kleiner Teil des rezeptiven Feldes betrachtet (z. B. ein 3×3-Raster). Fukushima gibt

in [8] 12 solcher grundlegenden Linienformen an, mit denen die Schicht U_{S1} trainiert wurde. Diese sind in Abbildung 1 dargestellt. Durch die nachfolgende Schicht U_{C1} werden die erkannten Formen „verschwommen“, was die Deformations- und Skalierungs-Toleranz des Neocognitrons ausmacht. Auf der nächsten Stufe werden primitive Formen kombiniert, um globalere Muster zu erkennen. Ein Beispiel gibt Abbildung 2 an. Die dort dargestellten Formen setzen sich aus primitiven Linienstrukturen der vorhergehenden Stufe zusammen. Je höher die Nummer einer Stufe im Neocognitron ist, desto globalere Formen werden durch die entsprechende S-Schicht erkannt und desto größer wird der Teil des von einer Zelle betrachteten rezeptiven Feldes, was dazu führt, dass Ebenen höherer Stufen weniger Zellen enthalten ([2]). Die letzte Schicht, eine C-Schicht, deckt schließlich das rezeptive Feld im Ganzen ab. Jedes hierin enthaltene Neuron entspricht einem gelernten Muster, seine Ausgabe ist ein Maß für die Wiedererkennung dieses Musters, wenn eine Eingabe am rezeptiven Feld durch das Neocognitron geleitet wurde. Somit dient die letzte C-Schicht als *Klassifizierungsschicht*, an deren Ausgaben die tatsächliche Erkennung des Netzes abgelesen werden kann. Nur durch diesen speziellen Grundaufbau aus abwechselnden Schichten von S- und C-Zellen wird eine effektive deformationstolerante Mustererkennung durch das Neocognitron erreicht. Der tatsächliche Aufbau wie z. B. die Anzahl von Schichten variiert dabei je nach Anwendungsgebiet, allgemein ist für komplexere Muster eine größere Anzahl von Schichten erforderlich.

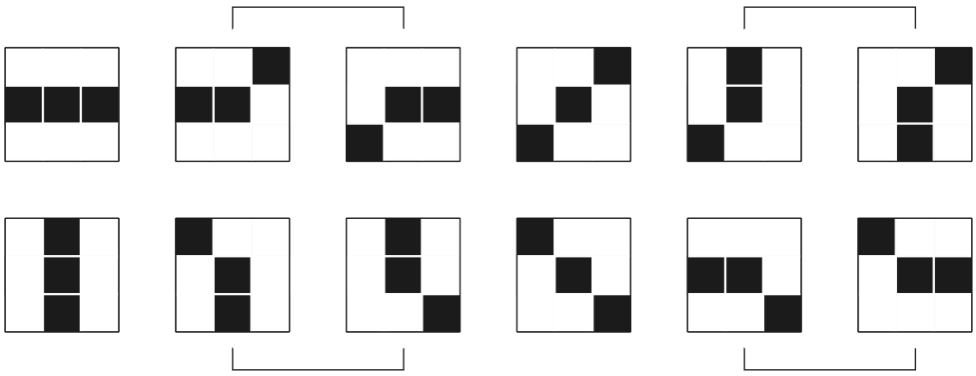

Abbildung 1: Primitive Muster, mit denen U_{S1} trainiert werden kann (nach [8])

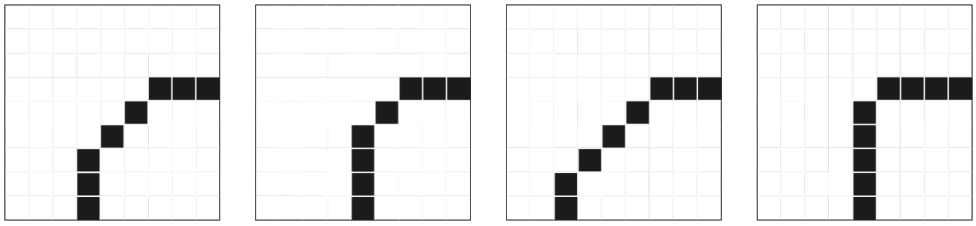

Abbildung 2: Beispiele von zum Training von U_{S2} verwendeter Muster (nach [10])

Abbildung 3 zeigt in vereinfachter Form die Struktur des Neocognitrons. Dabei ist pro Ebene einer S- bzw. C-Schicht jeweils nur eine Zelle dargestellt. Jede Zelle erhält ihre Eingabe von mehreren Zellen der Vorgängerschicht, was durch die kegelförmigen Muster zum Ausdruck kommt. Eine genauere Beschreibung dieser Verbindungen findet in den nachfolgenden Abschnitten statt. In

der Schicht U_{C3} besteht jede Ebene aus nur noch einem Neuron. Diese Schicht fungiert als Klassifizierungsschicht, ihre Ausgaben entsprechen dem Erkennungsgrad der gelernten Muster für die in U_0 anliegende Eingabe.

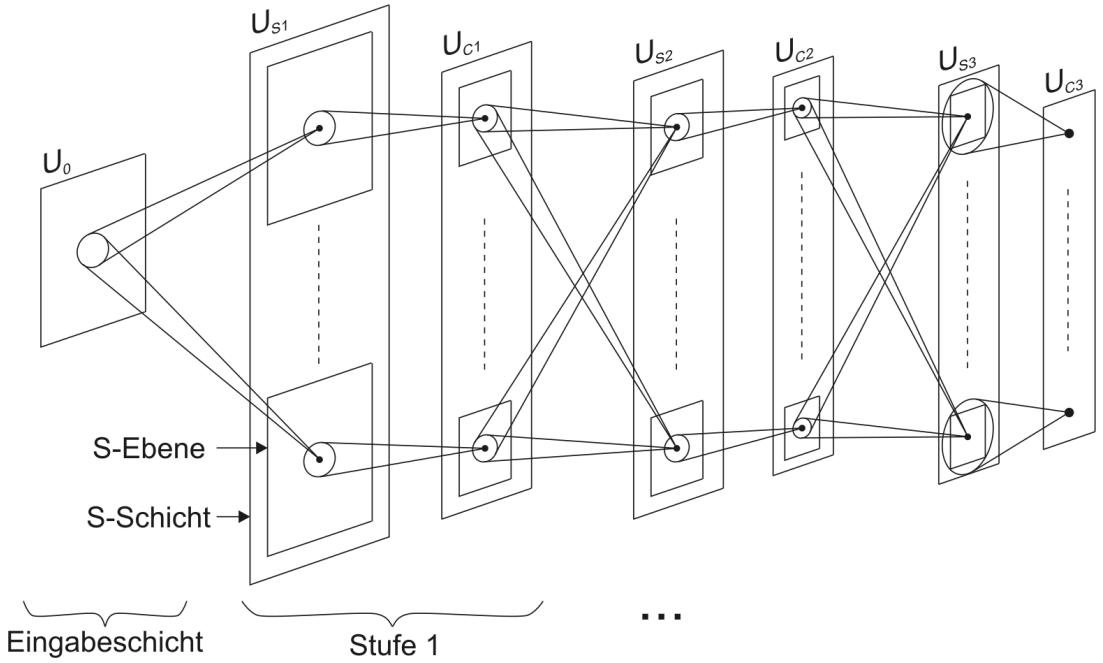

Abbildung 3: Struktur und grundlegende Verbindungsstrategie des Neocognitrons (nach [1])

2.2 Aufbau und Verbindungsstruktur der Schichten

2.2.1 S-Schichten

Jede S-Schicht ist vom prinzipiellen Aufbau und der Verbindungsstruktur von der vorherigen C-Schicht (bzw. Eingabeschicht) her gleich. Sie besteht aus mehreren Ebenen, welche rechteckige Felder von S-Zellen sind. Alle Ebenen einer S-Schicht haben dieselbe räumliche Ausdehnung, enthalten also dieselbe Anzahl von S-Zellen in gleicher Dimensionierung. Die Aufgabe einer S-Schicht besteht darin, Muster aus einem fest vorgegebenen Bereich der vorherigen Schicht zu extrahieren. Welche Muster dabei erkannt werden sollen, wird durch den Lernprozess des Neocognitrons festgelegt. I.d.R. gilt, dass jede Ebene einer S-Schicht genau ein Muster zu erkennen hat. Das bedeutet gleichermaßen, dass es pro S-Schicht genauso viele Ebenen gibt wie Muster, die durch die Schicht erkannt werden sollen. Jede S-Zelle ist in einer bestimmten Konfiguration mit einer Gruppe von Zellen der Vorgängerschicht verbunden. Allgemein wird ein Feld von $n \times n$ Zellen abgedeckt, von denen eine gewichtete Summe gebildet wird und als Eingabe der S-Zelle dient. Doch das Verbindungsmuster ist nicht auf quadratische Formen beschränkt, es kann sich dabei ganz im Gegenteil um beliebige Formen handeln (z. B. Rechtecke, Kreise, Dreiecke, ...). Alle S-Zellen einer Ebene, aber auch einer ganzen Schicht sind dabei in derselben räumlichen

Konfiguration mit Zellen der vorherigen Schicht verbunden. Benachbarte Zellen in einer Ebene stellen dabei eine i. Allg. um 1 Pixel parallel verschobene Gruppe von Zellen der Vorgängerschicht dar, sodass durch die Summe aller S-Zellen einer Ebene das gesamte rezeptive Feld abgedeckt wird. Weiterhin werden durch die Verwendung derselben räumlichen Konfiguration von S-Zellen einer Ebene für jede Ebene lediglich die Eingangs-Gewichte genau einer S-Zelle angelernt und abgespeichert. Alle anderen Zellen der Ebene „teilen“ sich die Gewichte mit dieser Zelle, womit alle diese Zellen auf dasselbe Muster reagieren, nur an unterschiedlichen Positionen im Eingabefeld. Nur durch dieses Prinzip des *weight sharing* sind die Netze des Neocognitrons überhaupt noch handhabbar ([16]).

Abbildung 4 zeigt beispielhaft die Belegung von S-Zellen einer Ebene der Schicht U_{S1} mit 3×3 Zellen der Eingabeschicht U_0. Die beiden markierten Zellen links oben in der $S1$-Ebene sind dabei benachbart, entsprechend ergibt sich eine Parallelverschiebung ihrer räumlichen Konfiguration in U_0. Die Zellen in den Ecken von U_{S1} bzw. an den Rändern des Feldes können keine 3×3 Zellen von U_0 abdecken und erhalten ihre Eingaben von entsprechend kleineren Gebieten.

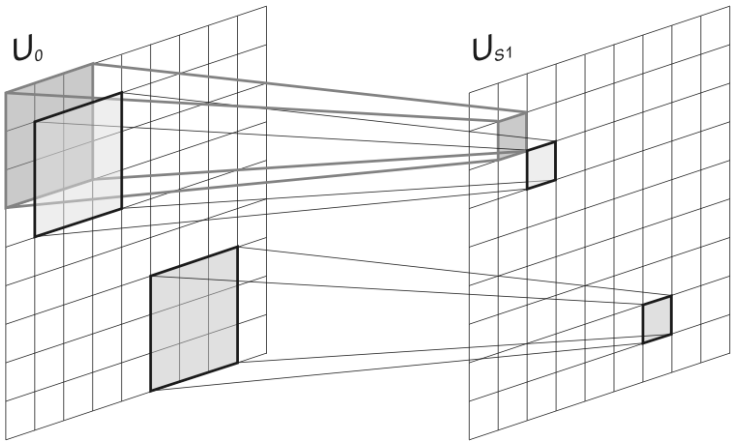

Abbildung 4: Beispielhafte Verbindungen von Zellen der Eingabeschicht U_0 mit U_{S1}

Durch die Abdeckung des gesamten rezeptiven Feldes mit einer Ebene wird wenigstens eine S-Zelle aktiviert, wenn das Muster, auf welches die Ebene angelernt wurde, im rezeptiven Feld an beliebiger Stelle auftritt. Die Verbindung zu S-Zellen einer Ebene ist daher dafür zuständig, dass Muster unabhängig von ihrer Position erkannt werden, wodurch die Eigenschaft der Positionsinvarianz erreicht wird. Abbildung 5 zeigt exemplarisch, wie ein Muster bestehend aus 3 vertikalen Pixeln zentriert in einem 3×3-Feld zu einer Aktivierung von S-Zellen genau der Ebene von Schicht U_{S1} führt, welche auf dieses Muster angelernt wurde. In der Mitte von U_0 tritt das Muster zweimal auf (überlappend durch die 4 vertikalen Pixel), zusätzlich noch einmal in der linken unteren Ecke. Es lässt sich gut erkennen, dass genau die S-Zellen der Ebene aus U_{S1} voll aktiviert werden, die mit dem Gebiet aus U_0 verbunden sind, welches das gelernte Muster komplett enthält. Kommt das Muster in einem Gebiet zum Teil vor, so werden die entsprechenden S-Zellen nur noch gering aktiviert, was durch die hellgrauen Zellen dargestellt ist. Stimmen zu wenig Pixel mit dem gelernten Muster überein, so sorgt eine Hemmung (durch die V-Zelle, siehe

unten) dafür, dass es zu keiner Aktivierung der S-Zelle kommt.

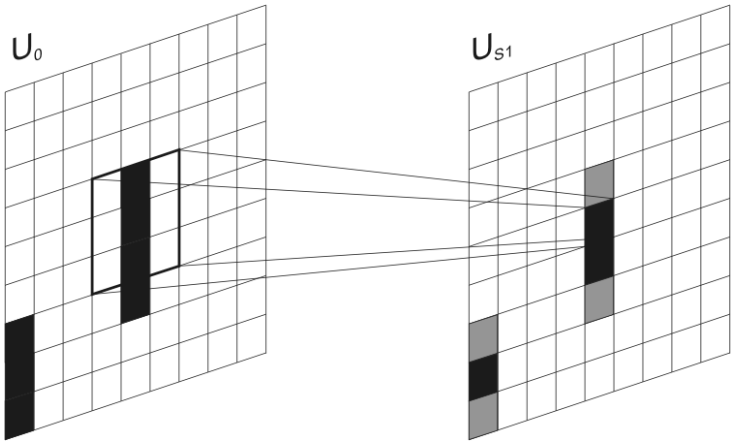

Abbildung 5: Beispiel für ein eingegebenes Muster und die Aktivierung von S-Zellen dadurch

S-Ebene höherer Schichten, die ihre Eingaben von C-Schichten anstelle und nicht direkt vom Eingabefeld erhalten, sind nicht mit nur einer Ebene der vorherigen C-Schicht verbunden, sondern mit *allen* darin enthaltenen Ebenen, wobei jeweils dasselbe Gebiet jeder C-Ebene überdeckt wird. Dadurch werden unterschiedliche Muster, die in der Vorgängerstufe erkannt wurden, zu globaleren Merkmalen zusammen gefasst und es kann darauf reagiert werden. Das Schema dieser Verbindungsstruktur wird auch in Abbildung 3 dargestellt.

Ein wichtiges Detail wurde bislang verschwiegen und soll an dieser Stelle ergänzt werden. Neben S- und C-Zellen gibt es noch einen dritten Zellentyp, die V-Zellen. Jeder S-Zelle ist eine V-Zelle zugeordnet, deren Ausgabe als inhibitorischer (hemmender) Faktor in die Eingabe der S-Zelle mit einfließt. Im ursprünglichen Neocognitron (siehe [1]) wurde auch jeder C-Zelle eine V-Zelle zugewiesen und es wurde damit zwischen V_S- und V_C-Zellen unterschieden, in späteren Veröffentlichungen (vgl. [2]) fehlt diese Unterscheidung, dort finden nur noch die zu S-Zellen zugehörigen V-Zellen Verwendung. Auf diese soll sich auch in dieser Arbeit beschränkt werden. Wie bereits erwähnt wirkt die Ausgabe einer V-Zelle hemmend auf die Eingabe der zugehörigen S-Zelle ein. Der Sinn hierfür ergibt sich aus der Verbindungsstruktur einer V-Zelle. Diese erhält ihre Eingaben von denselben Zellen wie ihre zugehörige S-Zelle. Abbildung 6 verdeutlicht diese Struktur. Das Gewicht von V- zu S-Zelle ist weiterhin so trainiert, dass die Hemmung der V-Zelle mit der Stärke des Eingabesignals variiert. Wird genau das gelernte Muster an die S-Zelle angelegt, so ist die Aktivierung dieser größer als die Hemmung ihrer V-Zelle, sodass das Muster wiedererkannt wird. Gibt es kleine Abweichungen, so wird die S-Zelle zwar immernoch aktiviert, doch der Einfluss der Hemmung nimmt zu. Gibt es hingegen nur noch kleine Übereinstimmungen, so ist die Hemmung hoch und die S-Zelle wird nicht aktiviert. Würde es kein inhibitorisches Signal geben, so wäre die Aktivierung der S-Zelle bei geringen Übereinstimmungen immernoch gegeben und das, obwohl nicht das von ihr gelernte Muster als Eingabe angelegt wurde. V-Zellen helfen S-Zellen somit zwischen relevanten und irrelevanten Strukturen zu unterscheiden und spielen in

dieser „Wächter“-Funktion eine wichtige Rolle bei der Mustererkennung (siehe [3]).

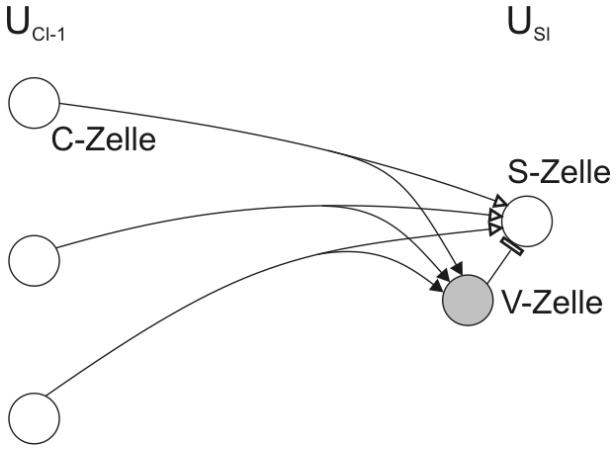

Abbildung 6: Verbindungen von V- und S-Zelle sowie Vorgängerschicht U_{Cl-1}

2.2.2 C-Schichten

Durch S-Schichten können kleinere Abweichungen vom gelernten Muster kompensiert werden und S-Zellen reagieren auf gelernte Muster unabhängig von ihrer Position im Eingabefeld, wobei Positionsinformationen erhalten bleiben. Durch S-Zellen lassen sich jedoch keine größeren Verformungen und Skalierungen kompensieren. Hier kommen die C-Schichten ins Spiel, wobei in einer Stufe des Neocognitrons auf jede S-Schicht eine C-Schicht folgt. C-Schichten haben einen ähnlichen Aufbau wie S-Schichten. Sie bestehen aus mehreren Ebenen, dies sind rechteckige Felder von C-Zellen. C-Zellen erhalten weiterhin ihre Eingabe ähnlich zu S-Zellen von einem festgelegten Gebiet aus Zellen der vorhergehenden S-Schicht. Alle C-Zellen einer Ebene decken dabei das gesamte rezeptive Feld ab und jede C-Zelle einer Schicht ist mit einem Gebiet derselben räumlichen Konfiguration der vorherigen S-Schicht verbunden. Benachbarte C-Zellen einer Ebene erhalten dabei ihre Eingaben von benachbarten parallel verschobenen Gebieten aus S-Zellen, auch dieses Prinzip ist mit S-Schichten äquivalent. Der Betrag der Verschiebung beträgt allgemein 1 oder 2 Zellen. Durch eine Parallelverschiebung um 2 S-Zellen ist es möglich die Dichte von C-Zellen im Vergleich zur vorhergehenden S-Schicht zu halbieren (siehe [7]) und somit auf das Ziel hinzuwirken, dass die Kategorieschicht nur noch eine Zelle pro Ebene enthält. C-Zellen sind somit fähig, die Inhalte von S-Zellen der vorhergehenden Schicht zu komprimieren.

In der Regel sind C-Zellen einer Ebene nur mit einer S-Ebene verbunden. Sie reagieren damit auf dieselbe Eigenschaft wie die entsprechende S-Ebene. Der Bereich von S-Zellen, mit denen eine C-Zelle dabei verbunden ist, entspricht somit der Erkennung des gleichen Musters an leicht versetzten Positionen. Eine C-Zelle reagiert also auch dann noch auf ein Muster, wenn es in der Eingabe leicht verschoben ist. Der Grad der Verschiebbarkeit variiert dabei mit der Größe des Gebiets verbundener S-Zellen. Ist das Gebiet klein, so werden nur kleine Verschiebungen

kompensiert, bei zu großen Gebieten kann es dazu kommen, dass zwei erkannte Muster an unterschiedlichen Positionen zu einem Gebiet verschmelzen, was einen Informationsverlust bedeutet. Wenn ein Muster im Eingaberaum an unterschiedlichen Positionen präsentiert wird, so variiert die Reaktion von C-Zellen in mittleren Schichten abhängig von der Position des Musters. Je höher eine Schicht im Neocognitron angesiedelt ist, desto geringer fällt diese Variation aus, auch aufgrund der Komprimierung von Informationen durch C-Schichten. Zellen der letzten C-Schicht (der Kategorieschicht) werden von der Verschiebung gar nicht mehr beeinflusst ([2]), was ein Verdienst der schrittweisen Kompensierung positionsbedingter Fehler mit Hilfe von C-Zellen ist. Diese Art der Tolerierung von Positions-Fehlern bei sequentiellem Durchlaufen aller Schichten anstelle des Versuchs, dies in nur einem Schritt zu tun, spielt eine wichtige Rolle sowohl bei der Positionsinvarianz als auch bei der Tolerierung von Verzerrungen durch das Netz ([3]).

Im Gegensatz zu S-Schichten gibt es für C-Schichten keine hemmenden Zellen, wodurch jede Ausgabe einer S-Zelle zu einer Stimulierung mit dieser Zelle verbundener C-Zellen führt. Durch die Überlappung von Gebieten, die mit C-Zellen verbunden sind und die Abdeckung des gesamten rezeptiven Feldes führt die Ausgabe einer S-Zelle zur Aktivierung mindestens einer C-Zelle, i. Allg. jedoch von mehreren benachbarten C-Zellen. Abbildung 7 verdeutlicht diesen Vorgang. Hierbei wird angenommen, dass die Gewichte in der Mitte des Verbindungs-Gebiets am größten sind und zum Rand hin abnehmen (siehe [2]). Es ist deutlich erkennbar, wie dadurch die in der Mitte des Gebiets befindliche C-Zelle stark aktiviert wird. Andere C-Zellen, welche mit der S-Zelle verbunden sind, wo diese aber nur am Gebiets-Rand auftritt, werden entsprechend geringer aktiviert.

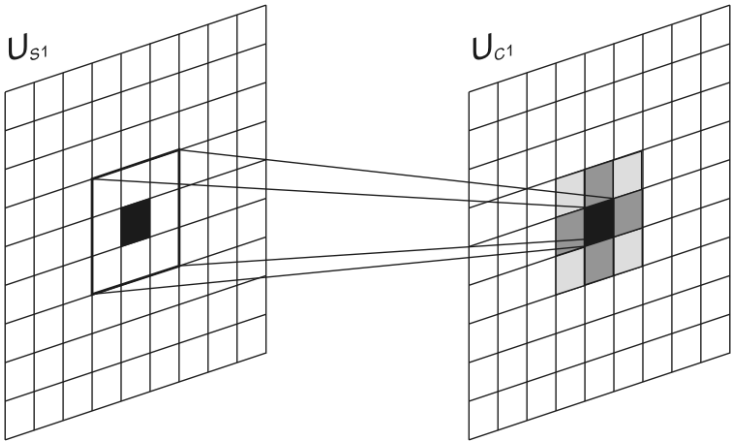

Abbildung 7: Aktivierung mehrerer C-Zellen durch eine S-Zelle

Aus Abbildung 7 heraus lassen sich weitere Rückschlüsse auf den Zweck von C-Zellen ziehen. Die Erkennung von S-Zellen wird „verschwommen“ ([3]), was in der nächsten S-Schicht dazu führt, dass deren Zellen auch dann noch aktiviert werden, wenn von ihnen gelernte Muster leicht verschoben oder teilweise deformiert sind, z. B. auch aufgrund von Skalierungen.

Ein Beispiel für die Verarbeitung eines Musters durch die erste Stufe des Neocognitrons wird

durch Abbildung 8 gegeben. Es wird angenommen, dass die dargestellte Ebene der Schicht U_{S1} das primitive vertikale Linienmuster bestehend aus 3 Pixeln gelernt hat. Wird das Muster in der Eingabe angelegt, so reagiert die S-Zelle am stärksten darauf, die mit dem Gebiet aus U_0 verbunden ist, welches das gelernte Muster beinhaltet. Eine geringe Aktivierung findet für die S-Zellen statt, welche das Muster noch zu zwei Drittel abdecken. Die anderen S-Zellen werden nicht aktiviert, auch wenn die Pixel vielleicht in ihrem Eingaberaum liegen, da die Hemmung der zugehörigen V-Zelle größer ist als die Anregung der S-Zelle. Die Ausgabe von U_{S1} wird nun an die erste C-Schicht U_{C1} weiter geleitet, von der hier eine Ebene dargestellt ist. Wie bereits beschrieben, werden die C-Zellen am stärksten aktiviert, für welche die Ausgabe der S-Zellen in der Mitte des Verbindungs-Gebietes liegt. C-Zellen, bei denen S-Zellen mit einer Ausgabe am Rand des Verbindungs-Gebietes liegen, werden geringer aktiviert (vorausgesetzt, die Gewichte von S- zu C-Zellen sind in der Gebiets-Mitte am größten und nehmen zum Rand hin ab). Der Effekt ist eine Ausdehnung der Aktivierung auf mehrere C-Zellen.

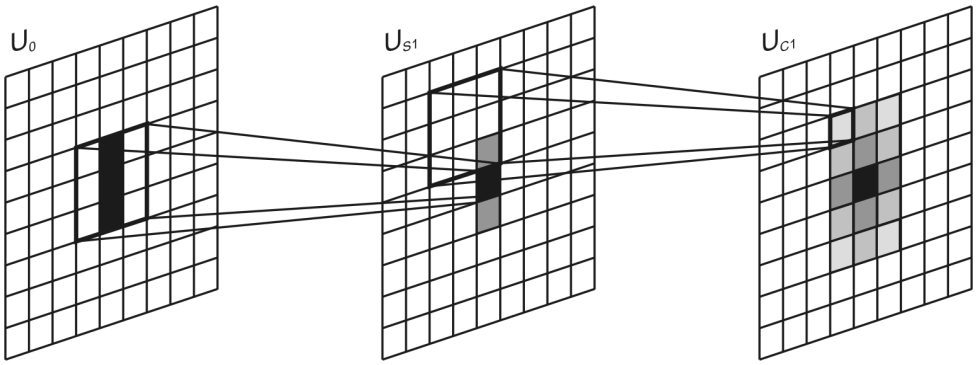

Abbildung 8: Verarbeitung eines gelernten Musters durch die erste Stufe des Netzes

Ein Wort noch einmal zur Verbindung von S- und C-Zellen. Meist wird eine C-Ebene mit einer S-Ebene der Vorgängerschicht verbunden, womit diese beiden Ebenen auf dasselbe Muster reagieren. Doch es gibt auch Ausnahmen hiervon. So gibt Fukushima in [7, 2, 8] Beispiele an, bei denen Muster, welche durch unterschiedliche Ebenen einer S-Schicht erkannt werden, durch die C-Schicht derselben Stufe zusammen gefasst werden. Das bedeutet, dass Muster, die vielleicht unterschiedlich aussehen, aber dasselbe Objekt beschreiben, zunächst durch speziell trainierte S-Zellen getrennt erkannt werden (da diese Fähigkeit bei größeren Unterschieden nicht von einer S-Zelle erbracht werden kann), um dann durch C-Zellen zur gleichen Objektgruppe zugeordnet zu werden. Damit wird versucht dem Umstand entgegen zu wirken, dass das Neocognitron auf Ähnlichkeiten zwischen Mustern arbeitet und nicht auf gebräuchlichen Schreibweisen von Zeichen. In Fukushima's einfachstem Beispiel hierfür (siehe [2, 8]) werden primitive Linien auf Stufe 1 des Neocognitrons zusammen gefasst. Abbildung 9 zeigt einen Auszug aus dieser Stufe. Die beiden unteren gelernten Muster werden als ähnlich angesehen, separat von U_{S1} erkannt und in U_{C1} in einer Ebene zusammen gefasst. Es bleibt zu erwähnen, dass solche Verdichtungen erkannter Muster problemabhängig festzulegen sind und nicht automatisiert werden können.

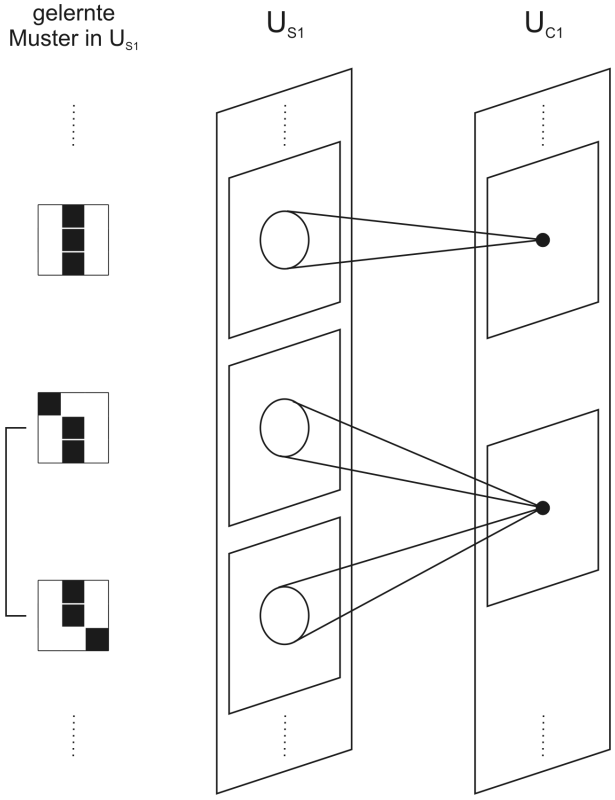

Abbildung 9: Zusammenfassung von S-Ebenen durch eine C-Ebene aufgrund ähnlicher Muster

2.3 Arten von Gewichten und Verbindungen

In der bisherigen Behandlung der Netzstruktur wurden zwar bereits Aussagen über die Verbindungen von Schichten miteinander getroffen, die Belegung dieser Verbindungen mit Gewichten kam dabei allerdings noch nicht zur Sprache. Doch genau in diesen Gewichten steckt wie bei den meisten anderen Netzmodellen auch das eigentliche Wissen des Neocognitrons, erst durch Anpassung dieser werden Muster gelernt und können schließlich auch wiedererkannt werden.

Grundlegend lässt sich zwischen zwei Gewichtsklassen unterscheiden: feste und variable Gewichte. Variable Gewichte lassen sich weiter unterteilen nach ihrer Zugehörigkeit zu exzitatorischen und inhibitorischen Verbindungen.

Gewichte a von der Eingabeschicht bzw. einer C-Schicht zu S-Zellen der nachfolgenden S-Schicht sind *variabel* und *exzitatorisch*, d. h. sie werden erst durch den Lernprozess festgelegt und bestimmen dann, welches Muster durch eine S-Ebene gelernt wurde. Ihr verstärkender Einfluss führt zur Aktivierung einer S-Zelle, wenn das gelernte Muster in deren Eingabegebiet angelegt wird. Durch das bereits beschriebene *weight sharing* müssen Eingangs-Gewichte für jede S-Ebene nur einmal abgespeichert werden und gelten dann für alle S-Zellen dieser Ebene.

Gewichte b von V-Zellen zu S-Zellen (eine V-Zelle ist mit genau einer S-Zelle verbunden, der sie zugehörig ist) sind *variabel* und *inhibitorisch*, d. h. dass auch sie erst durch den Lernprozess

festgelegt werden und dann mit der Ausgabe der V-Zelle hemmend auf die Eingabe der S-Zelle einwirken. Ihre Funktion besteht darin, S-Zellen relevante von irrelevanten Mustern unterscheiden zu lassen. Durch das *weight sharing* Prinzip muss auch hier nur ein Gewicht pro S-Ebene abgespeichert werden.

Gewichte c von der Eingabeschicht bzw. einer C-Schicht zu V-Zellen der nachfolgenden Schicht sind *fest* und *exzitatorisch*, d. h. sie werden vor dem Lernprozess bzw. bei der Konstruktion des Neocognitrons festgelegt und nicht mehr verändert. Sie führen dazu, dass eine V-Zelle aktiviert wird und hemmend auf ihre S-Zelle einwirkt, wobei die Hemmung bei Anliegen nicht gelernter Muster größer ist als die Anregung der S-Zelle. Eine Strategie für die Belegung der Eingangs-Gewichte einer V-Zelle besteht im Festlegen desselben Wertes für alle solche Gewichte, in [7] wird allerdings wie bei d-Gewichten eine zum Rand hin abnehmende Intensität vorgeschlagen (Gauss'scher Kernel). Durch das Prinzip des *weight sharing* müssen die Gewichte nur einmal für eine S-Ebene abgespeichert werden und gelten dann für alle anderen V-Zellen der Ebene. In der Regel teilen sich sogar alle V-Zellen einer gesamten S-Schicht dieselben Eingangs-Gewichte.

Gewichte d von einer S-Schicht zur nachfolgenden C-Schicht sind *fest* und *exzitatorisch*, d. h. sie werden vor dem Lernprozess z. B. bei der Konstruktion des Netzes festgelegt und können dann nicht mehr verändert werden. Durch sie kommt es (ähnlich zu einer Tiefpassfilterung in der Bildverarbeitung) zu einer Verteilung der Aktivierung einer S-Zelle auf mehrere mit ihr verbundene C-Zellen, womit kleine Verschiebungen und Deformationen kompensiert werden können. Die Strategie der Belegung dieser Gewichte kann unterschiedlich ausfallen. Beispielsweise können alle Gewichte eines Verbindungs-Gebietes auf denselben Wert gesetzt werden. Es wird jedoch vorgeschlagen (vgl. [2]), Gewichte in der Mitte eines Verbindungs-Gebietes auf einen hohen Wert zu setzen und zum Rand hin zu verringern. In [8] wird ein Gauss'scher Kernel angegeben mit $d(\nu) = \gamma_l^{|\nu|}$, wobei ν einen zweidimensionalen Vektor darstellt, der den Versatz zum Zentrum des Kernels angibt und $d(\nu)$ das Gewicht an dieser Stelle. γ_l wird für jede Stufe l einzeln gewählt und es gilt $0 < \gamma_l \leq 1$. Dies führt zu der höchsten Aktivierung, wenn S-Zellen mit hoher Ausgabe in der Mitte des Gebietes liegen sowie entsprechend geringerer Aktivierung am Rand. Abbildung 10 zeigt diese Struktur der Gewichtsmatrix an einem Beispiel für ein 5×5-Verbindungsgebiet.

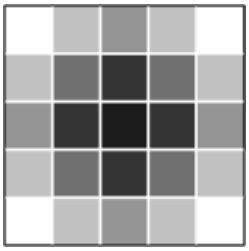

Abbildung 10: Beispiel für die Gewichtsmatrix zwischen S-Zellen und einer C-Zelle

Lovell *et. al.* [12] untersuchten den Gauss'schen Kernel von c- und d-Gewichten und stellten fest, dass dieser einen großen Einfluss auf die Erkennungsrate des Neocognitrons hat. So wurde

heraus gefunden, dass der Kernel weniger steil gewählt werden sollte, was in einer verbesserten Klassifizierung resultiert.

Zusammenfassend zeigt Abbildung 11 noch einmal alle Verbindungsarten auf.

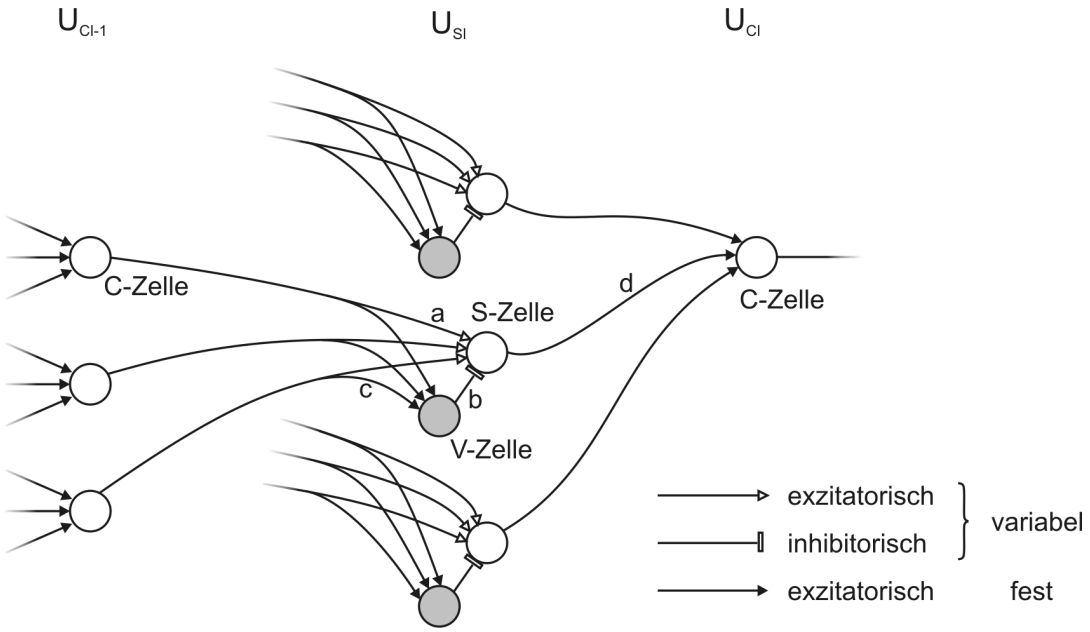

Abbildung 11: Übersicht der Verbindungsarten im Neocognitron (nach [8])

2.4 Mathematische Beschreibung

Nachfolgend soll die Verarbeitung durch das Neocognitron mathematisch beschrieben werden, auch um eine Implementierung zu ermöglichen. Die Beschreibung orientiert sich dabei an der vereinfachten, aber dennoch korrekten Darstellung in [16], um nicht die etwas umständliche Formulierung von Fukushima selbst (vgl. z. B. [8]) nutzen zu müssen. Diese bezieht die Verbindungsstruktur explizit mit ein, während nachfolgend einfach davon ausgegangen werden soll, dass es Verbindungen einer Zelle zu anderen Zellen gibt. Wie diese organisiert sind, wurde bereits mit dem Aufbau des Netzes in 2.2 beschrieben und soll hier ausgespart bleiben.

Jede S-Zelle sowie die zugehörige V-Zelle sind mit n Zellen der vorhergehenden C-Schicht (resp. der Eingabeschicht) verbunden. In die Berechnung der **Ausgabe y einer S-Zelle** fließen das von den verbundenen C-Zellen stammende *exzitatorische* Signal e sowie das von der V-Zelle ausgegebene *inhibitorische* Signal h mit ein, die wie folgt definiert sind:

$$e = \sum_i x_i a_i \qquad\qquad h = v \cdot b$$

Hierbei ist x_i die Ausgabe der i-ten verbundenen C-Zelle (resp. Eingabezelle), a_i die zugehörige variable Gewichtung, v die Ausgabe der zur S-Zelle gehörenden V-Zelle und b die von dieser Zelle

stammende variable Gewichtung. Die Ausgabe y einer S-Zelle ergibt sich dann wie folgt:

$$y = r \cdot \phi\left(\frac{1+e}{1+\frac{r}{1+r}\cdot h} - 1\right) \text{ mit } \phi(z) = \begin{cases} z, & \text{falls } z \geq 0 \\ 0, & \text{sonst} \end{cases}$$

Es gilt $r > 0$, $e \geq 0$ und $h \geq 0$ (alle Gewichte und Ausgaben sind stets ≥ 0) und somit ist $(1+e)/(1+\frac{r}{1+r}\cdot h) - 1 \geq 0$ genau dann, wenn $(1+e) \geq (1+\frac{r}{1+r}\cdot h)$, also $e \geq \frac{r}{1+r}\cdot h$. Das bedeutet, dass die Ausgabe einer S-Zelle größer 0 ist, wenn das exzitatorische Signal e größer dem gewichteten inhibitorischen Signal h ist. r bestimmt hierbei die *Selektivität* der S-Zelle. Durch diesen Wert wird festgelegt, welchen Einfluss das inhibitorische Signal hat. Für kleine r ist der Einfluss von h gering, mit steigendem r geht der Faktor $r/(1+r)$ gegen 1. Dies ist so interpretierbar, dass die S-Zelle bei geringer Selektivität auch auf solche Signale reagiert, die nur in geringem Maße dem von ihr gelernten Muster entsprechen, also eine starke Deformation aufweisen. Bei hoher Selektivität hingegen wird die S-Zelle nur aktiviert, wenn das angelegte Signal mit dem gelernten Muster gänzlich übereinstimmt. Durch r wird somit die Fähigkeit einer S-Zelle bestimmt, auf deformierte Muster zu reagieren. Die Wahl von r stellt dabei einen Spagat zwischen *Deformationstoleranz* und *Trennschärfe* zwischen unterschiedlichen Mustern dar. Mit dem Begriff der Trennschärfe wird die Fähigkeit des Neocognitrons bezeichnet, zwischen verschiedenen Mustern zu unterscheiden. Bei kleinem r ist die Deformationstoleranz hoch, die Trennschärfe jedoch gering, d. h. eine S-Zelle reagiert auch auf solche Muster, die vielleicht von anderen S-Zellen gelernt wurden und für die sie nicht zuständig sein sollte. Dies ergibt nach [4] eine Situation des gemeinschaftlichen Kodierens von Eigenschaften entgegen der Theorie von Großmutter-Zellen: viele Zellen einer S-Schicht reagieren auf dieselbe Eigenschaft. Ein hohes r erhöht die Trennschärfe, verringert aber die Befähigung zur Tolerierung kleinerer Deformationen (siehe auch [7]). Fukushima gibt in [8] unterschiedliche Selektivitäten r_l für jede S-Schicht l an. In seinem 4-stufigen Netz zur Erkennung handschriftlicher Zeichen verwendet er dabei die Werte $r_1 = 1.7$, $r_2 = 3.8$, $r_3 = 1.5$ und $r_4 = 1.0$. Es sei angemerkt, dass die Wahl von r nur manuell vor dem Lernvorgang festgelegt werden kann und nicht trivial ist. Gleichzeitig stellt die Selektivität allerdings einen Schlüsselfaktor für das Funktionieren des Gesamtsystems dar. Lovell *et. al.* untersuchten in [12] unter anderem den Einfluss der Selektivität und stellten eine Methode vor, wie sich in Bezug auf das Neocognitron aus [8] die Selektivität bestimmen lässt, was in einer verbesserten Klassifizierungsrate resultiert.

Die **Ausgabe v einer V-Zelle**, die mit denselben n Zellen der Vorgängerschicht verbunden ist wie ihre S-Zelle, ergibt sich aus der Wurzel der Summe von den gewichteten quadrierten Ausgaben x_i der Vorgängerzellen (quadratischer Mittelwert):

$$v = \sqrt{\sum_i x_i^2 c_i}$$

c_i gibt dabei die feste Gewichtung von der i-ten Vorgängerzelle an.

Schlussendlich fehlt noch die Beschreibung der **Ausgabe x einer C-Zelle**. Diese ist immer dann größer 0, wenn mindestens eine mit der C-Zelle verbundene S-Zelle eine Ausgabe erzeugt hat. Dabei sei angenommen, dass die C-Zelle mit m S-Zellen der Vorgängerschicht verbunden ist. x wird dann erzeugt durch:

$$x = \psi\left(\sum_j y_j d_j\right) \text{ mit } \psi(z) = \frac{\phi(z)}{1+\phi(z)} \text{ und } \phi(z) = \begin{cases} z, & \text{falls } z \geq 0 \\ 0, & \text{sonst} \end{cases}$$

y_j stellt dabei die Ausgabe der j-ten verbundenen S-Zelle dar, d_j ist die mit dieser S-Zelle assoziierte feste Gewichtung. Die Funktion $\psi(z)$ stellt den Grad der Sättigung der C-Zelle dar. Durch die angegebene Definition dieser Funktion wird klar, dass die Ausgabe einer C-Zelle immer im Bereich [0..1) liegt. In [12] wird für den Divisor der Funktion $\psi(z)$ eine verallgemeinerte Version $\beta+\phi(z)$ angegeben, wobei β die „Geschwindigkeit“ der Sättigung angibt. Es wird angemerkt, dass Experimente gezeigt haben, dass β direkt die Erkennungsgenauigkeit beeinflusst, eine genauere Untersuchung fehlt dort allerdings.

3 Prozess der Mustererkennung

In diesem Abschnitt soll der Vorgang der Mustererkennung noch einmal zusammen gefasst und anhand eines kleinen Beispiels verdeutlicht werden. Dabei sei als Voraussetzung ein 3-schichtiges vollständig angelerntes Neocognitron angenommen, d. h. jede S-Ebene wurde auf ein spezielles Teilmuster trainiert. Niedere S-Schichten werden mit primitiven Mustern angelernt, während höhere S-Schichten Muster erkennen können, welche aus einfacheren Formen zusammen gesetzt sind.

Als Beispiel soll der Buchstabe `A` erkannt werden. Dazu wird dieser gerastert und die so entstehenden Grauwerte werden an die Eingabeschicht (das rezeptive Feld) angelegt. Die Verarbeitung findet schichtenweise in Vorwärtsrichtung durch das Netz statt, d. h. die Signale der Eingabeschicht werden zunächst an die erste S-Schicht U_{S1} angelegt. Das Gesamtmuster wird dabei „zerlegt“, indem S-Zellen auf angelernte Teilmuster reagieren, wenn diese in ihrem Eingaberaum angelegt werden. Positionsinformationen bleiben dabei durch die Position von S-Zellen in den einzelnen Ebenen der S-Schicht erhalten. In Abhängigkeit der Selektivität r reagieren S-Zellen auch auf Muster, die nicht exakt den gelernten Eigenschaften entsprechen. Im Beispiel der Eingabe von `A` könnten einzelne Linien und Schnittmuster verschiedener Linien erkannt werden. Abbildung 12 zeigt diesen Sachverhalt. Hierbei sind allerdings nur einige markante Muster gekennzeichnet, die erkannt werden (beispielsweise das $\wedge$ an der Spitze des `A`), tatsächlich wird der ganze Eingaberaum von Verbindungen mit S-Zellen abgedeckt.

Als nächstes werden die in U_{S1} erkannten Muster von der Schicht U_{C1} verarbeitet, welche die Erkennung tolerant gegenüber kleinen Verschiebungen und Deformationen macht. Dazu sind C-Zellen mit einem kleinen Bereich der vorhergehenden S-Schicht verbunden. Sobald eine S-Zelle

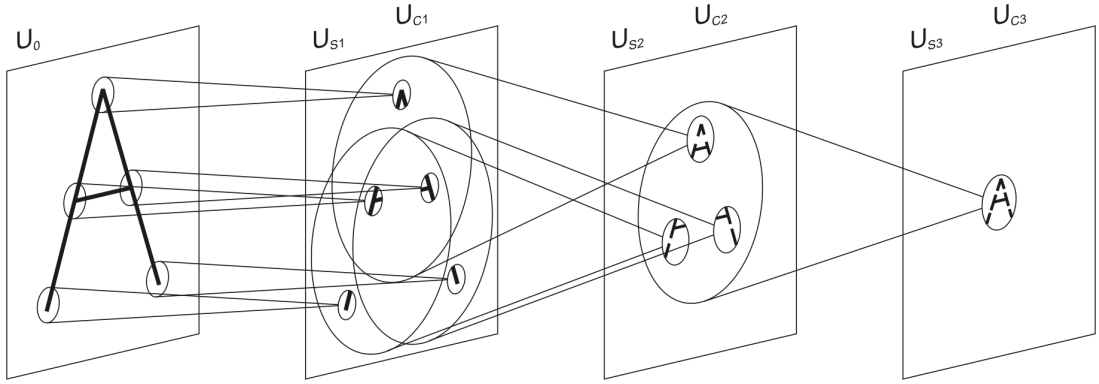

Abbildung 12: Beispiel für die Erkennung des Buchstaben `A` durch das Neocognitron (nach [8])

das von ihr gelernte Muster erkennt, reagieren auch alle mit ihr verbundenen C-Zellen darauf. In Abbildung 12 ist die Unterscheidung zwischen S- und C-Schicht zugunsten der Übersichtlichkeit nicht enthalten, dafür zeigt Abbildung 13, wie Zellen in U_{C1} einen kleinen Bereich von Zellen aus U_{S1} abdecken und so auf Muster reagieren, die an diesen Positionen erkannt werden.

Durch die nächste Schicht U_{S2} werden Eigenschaften erkannt, die sich aus einfachen Mustern zusammen setzen und somit Objekte höherer Ordnung bilden. Dadurch vergrößert sich der Bereich des rezeptiven Feldes, der durch diese Schicht abgedeckt wird. Abbildung 12 zeigt in der 1. Stufe 3 Bereiche, die sich gegenseitig überlappen und mit S-Zellen der 2. Schicht verbunden sind. In Abbildung 13 wird dieses Prinzip für eine Ebene aus U_{S2} noch einmal konkretisiert. Die dargestellte Zelle aus U_{S2} „beleuchtet“ ähnlich dem Kegel einer Taschenlampe einen Bereich von Zellen aus U_{C1}. Dabei fallen hier die 3 oberen erkannten Muster in den Kegel, die beiden unteren liegen außerhalb davon und fließen nicht mit in die Signalberechnung der S-Zelle ein. Die detektierten Muster aus der 1. Stufe werden kombiniert und im dargestellten Beispiel als der obere Teil des Buchstaben `A` erkannt. In der 3. Stufe schließlich wird das rezeptive Feld einer S-Zelle weiter vergrößert, sodass der gesamte Eingaberaum abgedeckt wird. U_{C3} gilt dabei als Klassifizierungsschicht, die höchste Ausgabe einer hier enthaltenen Zelle bestimmt das Muster, welches durch das Neocognitron erkannt wurde.

4 Erkennung verformter Muster, Verschiebungsunabhängigkeit

Die Erkennung deformierter Muster und die Tolerierung von Verschiebungen fußen auf zwei grundlegenden Mechanismen. Der erste ist die Selektivität r von S-Zellen. Mit ihr lässt sich bestimmen, bis zu welchem Grad ein eingegebenes Muster von einem gelernten Muster abweichen darf, um noch als dieses identifiziert zu werden. Das bedeutet, dass kleinere Deformationen kompensiert werden können.

Der zweite Mechanismus findet sich in der Funktionsweise der C-Zellen, durch die kleinere Verschiebungen und Skalierungen kompensiert werden. Ohne das Vorhandensein von C-Zellen wür-

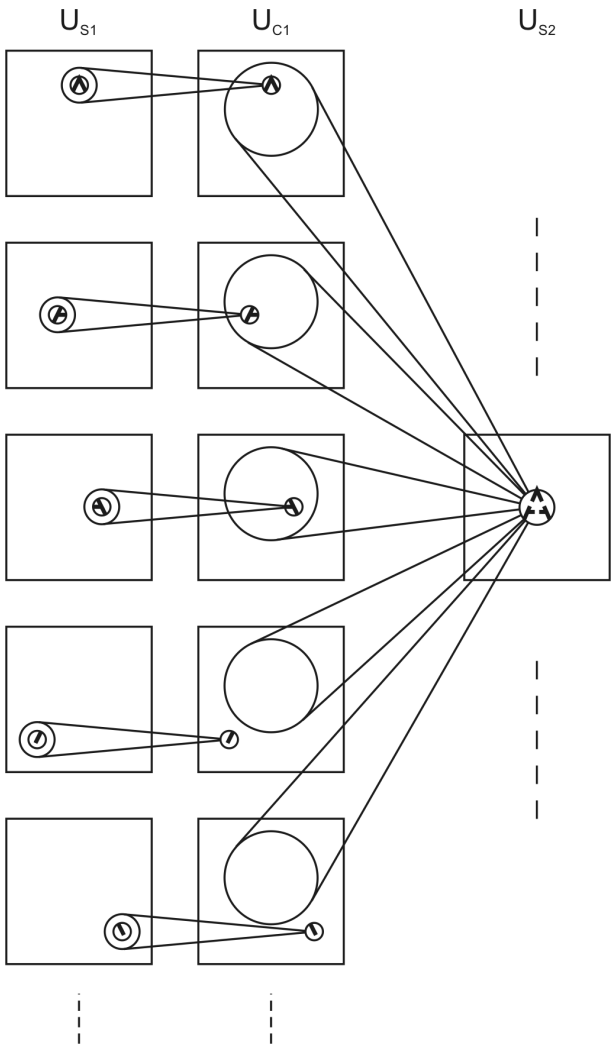

Abbildung 13: Erkennung von `A` über die 1. und 2. Stufe des Netzes (nach [8])

den S-Zellen nur dann auf ein Muster reagieren, wenn dessen Teilmuster an den beim Training vorgegebenen Positionen auftauchen. Im Beispiel von Abbildung 14(a) müssten die 3 dargestellten Teilmuster des Buchstaben `A` an eben diesen Positionen erscheinen. C-Zellen beziehen aber ein ganzes Gebiet von S-Zellen mit ein und so kommt es, dass sich die 3 Teilmuster an einer beliebigen Position in dem verbundenen Gebiet befinden dürfen und trotzdem erkannt werden, wie die Abbildungen 14(b) und 14(c) zeigen.

Problematisch wird der Einsatz von C-Zellen, wenn diese ein zu großes Gebiet an S-Zellen der Vorgängerschicht abdecken. In diesem Fall werden Muster selbst dann noch erkannt, wenn sie beispielsweise auseinander gerissen sind. Abbildung 15 zeigt diesen Sachverhalt. In 15(a) ist ein gelerntes Muster zu sehen, welches wiedererkannt wird. Eine Verschiebung wie in (b) ist unproblematisch, da alle S-Zellen einer Ebene auf dasselbe Muster trainiert sind und den Eingaberaum überlappend und vollständig abdecken. Anders als erwünscht wird aber auch das Muster in (c) als das gelernte erkannt, da es dieselben primitiven Eigenschaften besitzt, die von S-Zellen er-

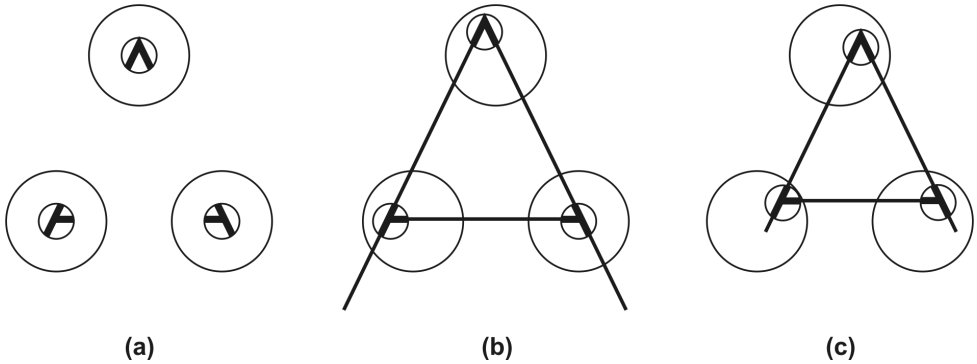

Abbildung 14: Funktionsweise der C-Zellen zur Erkennung deformierter Muster (nach [16])

kannt werden. Durch die nachfolgenden C-Zellen werden diese Eigenschaften auch dann noch detektiert, wenn sie wie in (c) gegeneinander verschoben sind, wobei trotzdem die Erkennung des gelernten Musters erhalten bleibt.

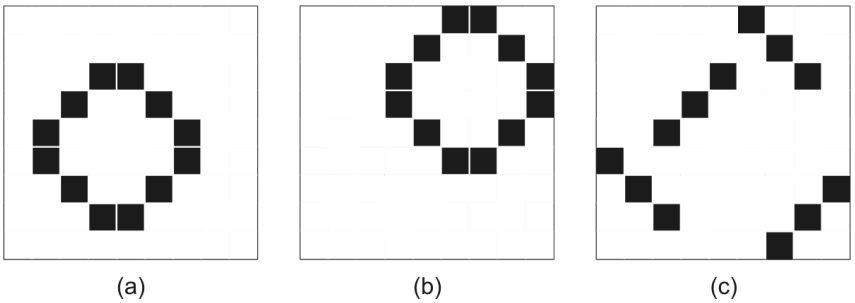

Abbildung 15: Probleme bei der Deformationserkennung (nach [10])

Insgesamt lassen sich zwei Dualitäten ausmachen. Zum einen die Dualität zwischen Deformationstoleranz und Trennschärfe bei Wahl der Selektivität r, was bereits in Abschnitt 2.4 zur Geltung kam. Zum anderen die Dualität zwischen Skalierungsunabhängigkeit und Trennschärfe bei der Wahl des Gebietes von S-Zellen, welches durch jeweils eine C-Zelle abgedeckt wird. Ist dieser Bereich klein, so werden zerrissene Muster wie das in Abbildung 15(c) abgelehnt. Allerdings können so keine Skalierungen kompensiert werden, durch welche Signale außerhalb des verbundenen Gebietes von S-Zellen liegen. Wurde der Bereich groß gewählt, so sind Skalierungen problemlos kompensierbar, allerdings werden unerwünschterweise auch Muster wiedererkannt, welche primitive Eigenschaften gelernter Muster enthalten, diese aber zerrissen sind. Eine Auflösung beider Dualitäten ist in praxi schwierig und muss von Fall zu Fall manuell erfolgen.

Zusammenfassend ist die Erkennungsleistung des Neocognitrons von vielen Faktoren abhängig. Darunter fällt die grundlegende Wahl der Anzahl von Schichten und der Trainingsmuster, das eingesetzte Lernverfahren (siehe Abschnitt 5) und eben auch die Wahl der Selektivitäten sowie des Gebietes von S-Zellen, mit dem eine C-Zelle verbunden sein soll. Insgesamt wird so ein Modell eines Neuronalen Netzes gekennzeichnet, welches durch seine stufenweise Tolerierung von Skalierungen und Deformationen gekennzeichnet ist, dafür aber eben auch eine langwierige manuelle

Anpassung der Parameter erfordert. Es existiert nur eine kleine Anzahl von Publikationen, welche sich mit der Erforschung des Neocognitrons und seines Verhaltens auseinander setzen. Lovell *et. al.* sehen in [11] den Grund darin in der Komplexität des Netzes, von demselben Autor existieren allerdings u. a. mit [11, 12] Veröffentlichungen, welche diesen Missstand durch eine Analyse der Parameter und des Netzes allgemein zu beheben versuchen.

5 Lernverfahren

Um Muster erkennen zu können, muss das Netzwerk einem Trainingsprozess unterworfen werden. Dieses Anlernen des Neocognitrons kann auf zwei unterschiedliche Arten erfolgen. Während Fukushima in [1] noch ein unüberwachtes Training vorschlägt, unterwirft er später in [8] ein Netz zur Erkennung handschriftlicher Zeichen einem überwachten Lernvorgang. Beide Varianten sollen im Nachfolgenden kurz vorgestellt werden, wobei das Hauptaugenmerk auf dem überwachten Lernen liegt. Ihnen ist gemein, dass der Lernvorgang in einer Anpassung der Eingabegewichte der S-Ebenen besteht und der Trainingsprozess schichtenweise von der ersten bis zur letzten Schicht erfolgt. Insgesamt ist festzuhalten, dass ein überwacht trainiertes Neocognitron gleich gut oder (i. Allg.) besser klassifiziert als ein unüberwacht angelerntes Netz derselben Dimensionierung.

5.1 Unüberwachtes Lernen

Das „Lernen ohne Lehrer“ ist biologisch am plausibelsten und wurde als erstes Lernverfahren des Neocognitrons von Fukushima 1980 in [1] vorgestellt. Dabei findet ein schichtenweises Lernen statt, d. h. eine höhere Schicht lernt erst dann, wenn der Lernvorgang der vorhergehenden Schichten komplett abgeschlossen wurde.

Anfänglich werden die exzitatorischen Eingangsgewichte der S-Zellen mit kleinen zufälligen positiven Werten belegt, sodass die S-Zellen sehr kleine unterschiedliche Orientierungen zeigen ([1]), die inhibitorischen Verbindungen von V-Zellen werden auf 0 gesetzt. Dann werden die anzulernenden Muster in beliebiger Reihenfolge präsentiert und das Netz wird mit ihnen angelernt. Der Lernvorgang einer S-Schicht gestaltet sich wie folgt: zunächst wird ein Trainingsmuster im Eingabefeld angelegt. Das Signal wird durch das Netz geleitet, bis es zu der aktuell anzulernenden Schicht gelangt. Da der Lernprozess vorhergehender Schichten bereits abgeschlossen ist, kann nun mit dieser Schicht fortgefahren werden. Als nächstes werden kleine Bereiche benachbarter S-Zellen über alle Ebenen der S-Schicht hinweg betrachtet. Aus solch einer „S-Spalte“ (S-column, siehe [1]) wird diejenige Zelle ausgewählt, welche die stärkste Ausgabe liefert. Diese wird als Kandidat für einen Repräsentanten ihrer Ebene gemerkt. Der Vorgang wird für alle S-Spalten wiederholt, wodurch man am Ende Ebenen erhält, die keinen, einen oder mehr als einen Kandidaten enthalten. Bei Ebenen, die mehr als einen Kandidaten enthalten, wird der als Repräsentant

der Ebene gewählt, welcher die höchste Ausgabe aufweist. Das ist gleichbedeutend mit der Auswahl genau des Musters, auf welches eine Zelle der Ebene am stärksten reagiert hat. Ebenen, die keine Kandidaten enthalten, haben zunächst auf kein angelegtes Muster als Gewinner reagiert und können durch den aktuellen Lernschritt nicht belegt werden. Besitzt eine Ebene einen Repräsentanten, so hat dieser am stärksten auf ein Eingabemuster reagiert. Dessen Verbindungen werden verstärkt, wobei der Betrag der Verstärkung proportional zur Ausgabe der Zellen ist, mit denen die S-Zelle in Verbindung steht. Die Adaption der exzitatorischen Gewichtungen a_i und der inhibitorischen Gewichtung b erfolgt dabei durch:

$$\Delta a_i = q \cdot a_i \cdot x_i$$

$$\Delta b = q \cdot v$$

a_i ist dabei die Gewichtung zur i-ten verbundenen Zelle der Vorgängerschicht, x_i stellt deren Ausgabe dar. Die inhibitorische Gewichtung b hängt von der Ausgabe v der V-Zelle ab, q beschreibt eine positive Lernrate, die den Grad der Spezialisierung auf ein Muster in einem Lernschritt festlegt.

Wurde eine S-Zelle nach diesen Regeln angelernt, so werden deren Eingangsgewichtungen für alle Zellen der zugehörigen Ebene übernommen. Dies ist gleichbedeutend mit einer Spezialisierung der Ebene auf ein angelegtes Muster, egal an welcher Stelle im Eingabefeld es vorkommt. Dieses Vorgehen stellt eine Form des Wettbewerbslernens (competitive learning) mit „winner takes all“ (oder nach [2]: „elite education“) Strategie dar, bei der alle Ebenen einer Schicht untereinander in Konkurrenz stehen. Dadurch spricht eine S-Zelle nach kurzem Training nur noch auf ein oder wenige Muster an und es ist sicher gestellt, dass nicht mehrere Ebenen auf dasselbe Muster reagieren. Der Prozess wird für alle Trainingsmuster wiederholt, wodurch jede Ebene nach geringer Wiederholungsrate auf eine eigene Mustereigenschaft reagiert.

Allgemein werden wiederholt alle Trainingszeichen vorgelegt, bis sich eine zufrieden stellende Klassifizierung einstellt. Das Problem dabei ist laut [10], dass der Zeitpunkt für das Trainingsende unklar ist und es keine Garantie für ein Ende des Trainings gibt. Meist wird das Training solange fortgesetzt, bis der Prozentsatz falsch klassifizierter Zeichen unter eine festgelegte Schranke fällt, keine Verbesserungen mehr erreicht werden oder die Rate korrekt klassifizierter Zeichen beginnt abzunehmen. Insgesamt stellt das unüberwachte Lernen eine komfortable Lernform dar, da keine zu lernenden Teilmuster angegeben werden müssen. Nachteilig ist die Ununterscheidbarkeit von Mustern, die sich sehr ähnlich sind, jedoch zu verschiedenen Klassen gehören ([2]).

5.2 Überwachtes Lernen

Auch wenn sich mit unüberwachtem Lernen erste Erfolge erzielen lassen konnten, so wurde doch schnell klar, dass eine getrennte Klassifizierung sehr ähnlicher Muster, die sich lediglich in kleinen Details voneinander unterscheiden, nur durch einen überwachten Trainingsprozess erreicht

werden konnte, da das Neocogitron mit Ähnlichkeiten und nicht aufgrund von Konventionen arbeitet (siehe [2]). Geht man bei unüberwachtem Lernen vom Standpunkt einer biologienahen Modellierung aus, entspricht der überwachte Lernprozess der Sichtweise eines Ingenieurs, für den der Erfolg bei der schlussendlichen Klassifizierungsaufgabe entscheidend ist ([7]).

Der grundlegende Lernvorgang gestaltet sich wie beim unüberwachten Lernen: es findet ein schichtenweiser Lernprozess statt, von der niedrigsten zur höchsten Schicht.

Zu Beginn des Trainings werden alle anpassbaren Gewichtungen (Eingangsgewichte der S-Zellen) auf 0 gesetzt - eine kleine unterschiedliche Ausrichtung wie beim unüberwachten Lernen ist nicht notwendig. Da es sich beim überwachten Lernen um ein „Lernen mit Lehrer“ handelt, präsentiert ein Lehrer im Nachfolgenden die Trainingsmuster, welche durch das Neocognitron zu lernen sind. In einer anzulernenden Schicht wählt der Lehrer hierbei zunächst eine Ebene aus und dann das Muster, welches von der Ebene erkannt werden soll. I.d.R. wird dabei jeder Ebene ein anderes Muster präsentiert, manchmal wird eine einzige Ebene jedoch auch mit mehr als einem Muster antrainiert, was die Fähigkeit deformierte Muster zu erkennen erhöhen kann ([8]). Diese Art des Trainings wird auch „vollständig überwachtes Training“ ([16]) genannt. Im Gegensatz zum normalen überwachten Training, wo nur Eingaben und vom Netz erwartete Ausgaben präsentiert werden müssen, sind beim vollständig überwachten Training auch alle zu lernenden Teilmuster der verdeckten Neuronen anzugeben, was die Erstellung einer Trainingsmenge an Mustern zu einem aufwändigen als auch kritischen Prozess werden lässt. So hängt der Erfolg der Erkennung deformierter Muster hochgradig von der Wahl der lokalen Eigenschaften ab, die in den mittleren Schichten des hierarchischen Netzwerks extrahiert werden ([8]).

Hat der Lehrer eine S-Ebene und das von ihr anzulernende Teilmuster ausgewählt, so gibt er zusätzlich die Zelle der Ebene an, welche auf das präsentierte Muster mit einer Gewichtsänderung reagieren soll. Nur die Verbindungen dieser „Keimzelle“ werden angepasst. Dazu wird das zu lernende Muster im Eingabefeld angelegt und bis zur aktuell lernenden Schicht durch das Netz propagiert. Die davor liegenden Schichten wurden bereits angelernt und reagieren somit korrekt auf die Eingabe. Die Eingabe der Keimzelle wird für deren Lernvorgang verwendet, wobei der Betrag der Verstärkung proportional zur Intensität der Reaktion der Zelle ist, von der die verstärkte Verbindung ausgeht. Insgesamt werden die exzitatorischen Verbindungen und die inhibitorische Verbindung zur S-Keimzelle wie beim unüberwachten Lernen nach den folgenden Regeln adaptiert:

$$\Delta a_i = q \cdot a_i \cdot x_i$$

$$\Delta b = q \cdot v$$

Dabei stellt a_i die Gewichtung zur i-ten verbundenen Zelle der Vorgängerschicht dar und x_i beschreibt deren Ausgabe, b ist die inhibitorische Gewichtung, v die Ausgabe der V-Zelle und q eine positive Lernrate. In [8] wird eine Lernrate von 10^4 vorgeschlagen, der tatsächliche Wert spielt aber eine untergeordnete Rolle, er sollte nur genügend groß sein, damit Muster in einem

Trainingsschritt angelernt werden können. q sollte für alle Ebenen einer Schicht gleich groß sein, empfehlenswert ist ein einheitliches q für das Antrainieren des ganzen Netzes.

Durch diese Anpassung der Gewichte eignet sich die Keimzelle die Fähigkeit an, bei Präsentation der angelernten Eigenschaft im Eingabefeld auf diese zu reagieren. Die anderen Zellen der Ebene erhalten dieselben Eingabegewichtungen wie die Keimzelle und reagieren somit auf dasselbe Muster, wenn dieses an einer anderen Position im Eingabefeld präsentiert wird.

Dieser Vorgang wird für alle Ebenen der Schicht und für alle Schichten mit jeweils unterschiedlichen Mustern wiederholt. Dabei genügt es, ein Muster lediglich einmal zu präsentieren. Insgesamt erhält man im Gegensatz zum unüberwachten Lernen eine bessere Klassifizierung unterschiedlicher Muster und im Vergleich zu anderen Netzmodellen ist die Trainingszeit sehr kurz. Nach [13] wurden für das Anlernen eines Neocognitrons zur Erkennung handschriftlicher Zeichen auf einer SUN-SPARC-Station 13 Minuten benötigt, während für ein Feedforward-Netz unter Verwendung des Backpropagation-Algorithmus für dieselbe Aufgabe 3 Tage Trainingszeit notwendig waren. Diesen positiven Eigenschaften steht gegenüber, dass ein großer Aufwand bei der Erzeugung der Trainingsdaten betrieben werden muss, da diese für jede Schicht gesondert auszuwählen sind.

6 Modifikationen des Neocognitrons

Im Folgenden wird knapp auf die wichtigsten Modifikationen des Neocognitrons eingegangen, die zum einen eine Erkennung mehrerer Zeichen in einem Bild und zum anderen eine Verbesserung der Klassifizierung zur Folge haben.

6.1 Neocognitron mit *Selective-Attention*-Mechanismus

In [16] wird eine Erweiterung des Neocognitron-Netzes beschrieben, die von Fukushima selbst entwickelt wurde und bereits abgewandelt im Cognitron zum Einsatz kam. Bei dieser Erweiterung handelt es sich um den sogenannten *Selective-Attention-Mechanismus*. Hierbei werden Rückkopplungen eingeführt, wodurch das Netzwerk in die Lage versetzt wird seine „Aufmerksamkeit“ auf einzelne Muster im Eingabefeld zu richten, falls dieses mehrere Muster enthält. Dadurch ist es möglich, automatisch nacheinander alle im Eingabefeld vorkommenden Muster zu erkennen, ebenso können unvollständige Muster vervollständigt werden, bei gleichzeitiger Entfernung von Rauschanteilen.

Hierzu existieren lokale Rückkopplungen jeder Stufe des Netzes an die vorhergehende Stufe. Dadurch werden Zellen unterdrückt, die nur schwach erregt sind, wodurch nur die am stärksten aktivierten Zellen ihre Ausgabe weiterleiten können. Ebenso führte Fukushima neue Zelltypen ein, womit insgesamt die Komplexität des Netzes weiter anwächst und dadurch noch höhere Anforderungen an Rechenleistung und Speicherkapazität stellt.

Abbildung 16 zeigt Beispiele für erkannte Muster. Die jeweils obere Zeile stellt dabei die durch die Aufmerksamkeitssteuerung gefilterten Bilder dar, die untere Zeile zeigt die Maske, welche durch die Aufmerksamkeitssteuerung entsteht.

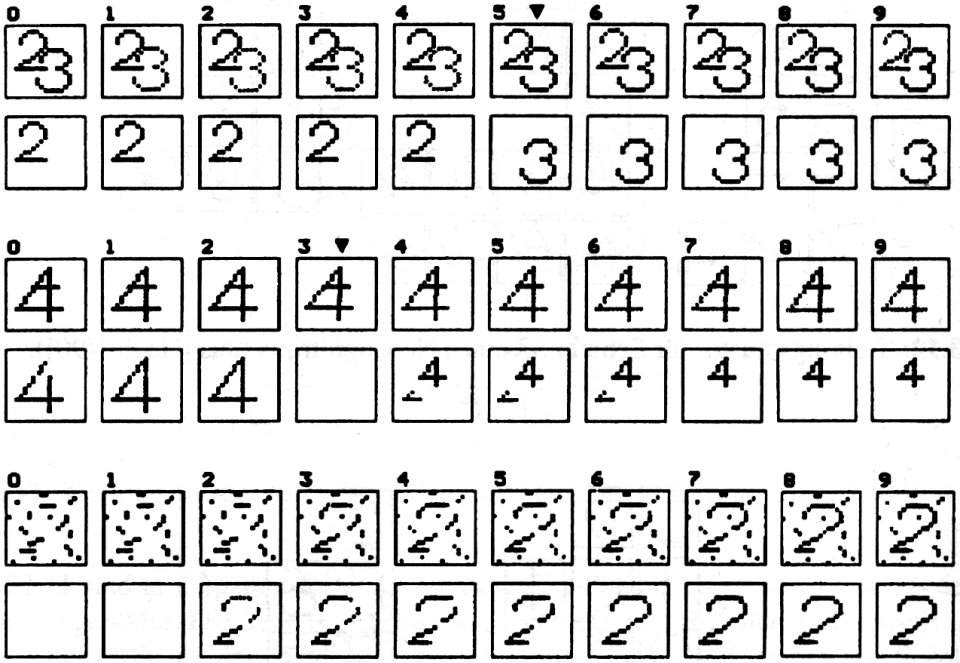

Abbildung 16: Beispiele für die Erkennung mittels *selective attention* (aus [16])

6.2 Doppelte C-Zellenschicht

Wie in [10] dargestellt wird, experimentierte Fukushima mit der Verdopplung der ersten C-Zellen-Schicht und stellte seine Ergebnisse 1994 vor. Grundlegend ist hierbei die Idee, dass deformierte Muster im Vergleich zu gelernten Mustern vor allem in den Randbereichen des rezeptiven Feldes große Abweichungen aufweisen. Die C-Schichten sind so verdrahtet, dass eine C-Schicht kleine Deformationen toleriert und die andere große. Eine Ebene der nachfolgenden S-Schicht ist dann so mit den C-Schichten verbunden, dass in ihrem Zentrum Verbindungen zu der C-Schicht für kleine Deformationen und an ihrem Rand Verbindungen zu der C-Schicht für größere Deformationen existieren. Die Fehlerrate konnte dadurch in praktischen Tests um bis zu 50% gesenkt werden. Doch auch diesem Verfahren lassen sich Nachteile zuordnen. Betrachtet man beispielsweise die Buchstaben `O` und `Q`, so unterscheiden sich diese nur gering in ihren Randbereichen und können dadurch mit der groben Deformationstolerierung durch C-Zellen in diesem Gebiet nicht unterschieden werden.

6.3 Rotations-invariantes Neocognitron

Das usprüngliche Neocognitron ist tolerant gegenüber Skalierungen, Verschiebungen und Verformungen von eingegebenen Mustern. Wichtig ist in einem Erkennungssystem jedoch auch die Unempfindlichkeit gegenüber Rotationen und dies kann das Neocognitron nicht leisten. Als möglichen Ausweg stellen Satoh *et. al.* in [14] eine Erweiterung vor, ein *rotations-invariantes Neocognitron*. In diesem besteht eine Schicht aus Stapeln von Ebenen, jeder Stapel ist für die Erkennung eines spezifischen Musters in unterschiedlichen Rotationen verantwortlich. Ein Stapel besteht dann aus mehreren Ebenen, wobei jede Ebene für einen bestimmten Grad an Rotation zuständig ist. Dabei erfolgt eine Abstufung in 30°-Schritten. Wird nun ein rotiertes Muster eingegeben, so kommt es zu einer Erregung derselben Stapel, die auch auf das Muster ansprechen, wenn es nicht rotiert ist. Es werden allerdings andere Ebenen innerhalb des Stapels aktiviert, welche dem Rotationswinkel entsprechen. Die in [14] enthaltenen Beispiele zeigen die Möglichkeiten des Systems auf und die Autoren zeigen sich zuversichtlich, dass es effektiv bei der Handhabung von Daten aus der realen Welt eingesetzt werden kann.

6.4 Verbessertes Neocognitron

Fukushima selbst stellt in [4] ein in mehreren Aspekten modifiziertes Neocognitron zur Erkennung handschriftlicher Ziffern vor, welches zum einen ein verbessertes Modell des visuellen Wahrnehmungssystems darstellt und zum anderen eine höhere Erkennungsrate liefert. Die Struktur des ursprünglichen Neocognitrons wird zunächst zu einem großen Teil beibehalten mit dem Unterschied, dass zwischen den Schichten U_0 und U_{S1} eine kontrast-extrahierende Schicht U_G eingeführt wird. Zellen hierin entsprechen retinalen Ganglion-Zellen, welche im visuellen System eben der Erkennung lokaler Kontraste dienen. U_G enthält 2 Ebenen, die jeweils die Extraktion positiver respektive negativer lokaler Kontraste (entsprechend On-Center und Off-Center Ganglion-Zellen) übernehmen. Eine weitere größere Modifikation stellt die Einführung inhibitorischer Verbindungen in der Umgebung von exzitatorischen Verbindungen zu C-Zellen der Schichten U_{C1} und U_{C2} dar. Jede Verbindung einer S- zu einer C-Zelle erhält dabei eine konzentrische inhibitorische „Umrahmung“. Dadurch erhält die C-Zelle die Charakteristik auf End-Linienstücke sowie Krümmungspunkte stärker zu reagieren als auf mittlere Linienpunkte, womit sich C-Zellen wie hyperkomplexe Zellen im visuellen Kortex verhalten. Endpunkte und Krümmungspunkte von Linien stellen markante Eigenschaften dar, auf die so verstärkt reagiert werden kann. Ein weiteres Resultat dieser lokalen Inhibition ist das Entstehen einer inaktiven Trennzone zwischen 2 unterschiedlichen Signalen, die nahe beieinander liegen und sonst verklumpen würden. Dadurch wird es S-Zellen der nächsten Schicht ermöglicht, in einfacher Weise die Anzahl ursprünglicher Merkmale zu erkennen, selbst nach der Unschärfe-Operation von C-Zellen. Als weiteren Schritt veränderte Fukushima das Lernverfahren des Neocognitrons. U_{S1} (die kantenextrahierende Schicht) wird überwacht trainiert, die mittleren Schichten U_{S2} und U_{S3} werden unüberwacht angelernt, ähnlich

der Methode, die bereits in Abschnitt 5.1 dieser Arbeit vorgestellt wurde. S-Zellen der höchsten Schicht U_{S4} unterliegen hingegen einem überwachten Wettbewerbs-Lernen, in welchem ein Lehrer dafür Sorge trägt, dass Muster korrekt klassifiziert werden.

In Computersimulationen auf der ETL1-Datenbank (einem Datensatz für segmentierte handschriftliche Zeichen) zeigte das Netz in einem Blindtest über 3000 Zeichen eine Erkennungsrate von 98.6%, auf dem Trainingssatz wurden sogar 100% erreicht. Im letzten Abschnitt von [4] bezieht Fukushima zu Aussagen anderer Stellung, dass das Neocognitron ein komplexes Netzwerk sei. Er hält dies für falsch. Operationen zwischen Zell-Ebenen können als zweidimensionale Filteroperationen aufgefasst werden, wobei der Filter aufgrund des „weight sharing“ für alle Zellen einer Ebene derselbe ist. Im Vergleich der Anzahl der verschiedenen Operationen, die so ausgeführt werden müssen, ist das Neocognitron Fukushima's Aussage nach ein sehr einfaches Netzwerk im Vergleich zu anderen KNNs.

7 Anwendungen

In diesem Abschnitt werden kurz ein paar Anwendungsgebiete des Neocognitrons genannt, wie sie auch in der Literatur zu finden sind.

7.1 Erkennen von Drahtmodell-Objekten

Laut [13] benutzten Lee und Patterson 1991 ein Neocognitron, das im Stande war Büroobjekte anhand ihrer Drahtgittermodelle zu erkennen. Solche Objekte sind dadurch gekennzeichnet, dass sie sich aus Flächen zusammensetzen, die durch gerade Linien und Eckpunkte miteinander in Verbindung stehen, z. B. Tische, Karteikästen etc. . Es wird erwähnt, dass die Erkennung aufgrund der Verschiebe- und Verformungstolerierung des Neocognitrons „gute“ Ergebnisse liefert, ein Gütewert wird jedoch nicht angegeben.

7.2 Objekterkennung

In [13] wird ein Beispiel für die Erkennung von Objekten durch das Neocognitron angegeben. Danach trainierten Menon und Heinemann (1988) ein vierschichtiges Neocognitron so, dass es Militärfahrzeuge (Jeeps, Panzer etc.) erkennen konnte. Das Netz war in der Lage, Objekte selbst dann noch korrekt zu klassifizieren, wenn ein Bild einen Störanteil von bis zu 30% beinhaltete.

7.3 Erkennung handschriftlicher Zeichen

Nach [10] integrierten Fukushima und Imagawa 1993 in ein Neocognitron mit Rückwärtsverbindungen eine Suchsteuerung, um die Anzahl der Zeichen einzuschränken, die gleichzeitig betrachtet werden. Die Suchsteuerung führt ein Suchfeld horizontal über die zu erkennenden Buchstaben, stets in Richtung der größten S-Zellen-Aktivität. Das Netz ist dabei aufgrund der selektiven Wahrnehmung in der Lage auch mehrere Zeichen zu erkennen, die sich im Suchfeld befinden. So lassen sich Buchstaben erkennen, auch wenn sie miteinander verbunden sind.

7.4 Erkennung koreanischer Schiftzeichen

[10] gibt als Anwendungsgebiet des Neocognitrons die Erkennung koreanischer Schriftzeichen an, die sich aus sogenannten Graphemen zusammensetzen und nicht miteinander verbunden sind. Danach benutzten Kim und Lee 1991 ein Neocognitron mit Rückwärtsverbindungen, welches nach einem überwachten Trainingsprozess 24 Grapheme erkennen konnte. Ein Zeichen konnte dann dadurch erkannt werden, dass die einzelnen Grapheme extrahiert und schließlich aufgrund ihrer Lage zu dem eigentlichen Zeichen zusammengesetzt wurden. Tests zufolge konnte das Netz rund 79% der Testzeichen korrekt klassifizieren.

7.5 Extraktion von Symmetrieachsen

Fukushima stellt in [6] das Modell eines Neuronalen Netzes vor, dessen Architektur die unteren Stufen des Neocognitrons einbezieht und in der Lage ist, Symmetrieachsen zu extrahieren. Dazu gibt es mehrere Arten von Schichten: die Eingabeschicht U_0 wird gefolgt von einer kontrastextrahierenden Schicht U_G, welche die On- und Off-Ganglion-Zellen des menschlichen Auges imitiert, die positive und negative Kontraste erkennen. Im Anschluss daran lassen sich eine Neocognitron-typische S- und C-Schicht U_S und U_C erkennen, wobei die S-Schicht mit 32 Ebenen zur Extraktion primitiver Kanten unterschiedlicher Orientierung zuständig ist. Die C-Schicht ist in 3 Unterschichten aufgeteilt, welche jeweils einen unterschiedlichen Grad an Unschärfe darstellen. Jede Unterschicht enthält abermals 32 Ebenen, wobei jede Ebene einer C-Unterschicht mit genau einer Ebene von U_S verbunden ist. Auf U_C folgt eine Schicht U_H, deren Ebenen einzelne Symmetrie-Achsen erkennen und eine Schicht U_X, welche die Ausgaben von U_H integriert und damit die Symmetrie-Informationen als Ganzes liefert. In den in der Arbeit vorgestellten Ergebnissen werden Symmetrieachsen zuverlässig extrahiert, selbst bei komplizierteren Mustern wie Gesichtern, die einen kleinen Grad an Asymmetrie aufweisen.

7.6 Wiederherstellung und Erkennung teilweise verdeckter Muster

Fukushima schlägt in [5] ein Modell für ein Neuronales Netz basierend auf dem Neocognitron vor, welches in der Lage ist teilweise verdeckte Muster wiederherzustellen. Dazu werden rekurrente Pfade benutzt, wobei unverdeckte Muster hauptsächlich von Signalen aus niedrigen Schichten erkannt werden, während höhere Schichten des Netzwerks über Feedback-Signale verdeckte Muster rekonstruieren. Eine spezielle Schicht U_M erkennt dabei verdeckende Objekte und entfernt diese. Dieses neue Modell ermöglicht es verdeckte Muster selbst dann zu erkennen, wenn sie zu einem Teil deformiert sind und dem Netz in der Trainingsphase präsentiert wurden. Unbekannte Muster werden durch Interpolation und Extrapolation sichtbarer Kanten ebenfalls versucht zu rekonstruieren, ebenso werden Ähnlichkeiten zu bereits gelernten Mustern mit einbezogen. Die veröffentlichten Ergebnisse der Computersimulation sind vielversprechend und zeigen die Möglichkeiten dieses komplexen Netzwerks. Eine Einschränkung stellt die Unterscheidung zwischen verdeckten und verdeckenden Objekten dar, die auf Grauwertunterschieden beruht. Fukushima selbst bezeichnet dieses Problem der guten Segmentierung verdeckter Objekte als Gegenstand zukünftiger Untersuchungen.

8 Zusammenfassung und Ausblick

Die vorliegende Arbeit hatte zum Ziel das Neocognitron zu beschreiben, ein hierarchisches mehrschichtiges Netzwerk, welches für Zwecke der Mustererkennung konzipiert wurde. Zunächst wurde der allgemeine Aufbau beschrieben und die Unterteilung in einzelne Schichten beleuchtet, auf deren Funktion eingegangen wurde. Es wurde dargelegt, wie das Neocognitron in der Lage ist, Verschiebungen, Skalierungen und Deformationen zu tolerieren. Weiterhin wurde der Prozess der Mustererkennung und der Erkennung verformter Zeichen beschrieben. Mit dem unüberwachten und überwachten Lernen wurden die beiden hauptsächlich eingesetzten Lernverfahren des Neocognitrons erläutert, bevor in den letzten Abschnitten auf einige Modifikationen und Anwendungen des Netzwerks eingegangen wurde.

Zusammenfassend lässt sich sagen, dass es sich beim Neocognitron um ein künstliches neuronales Netz mit außergewöhnlichen Fähigkeiten zur Mustererkennung handelt, welche sowohl die Toleranz gegenüber Verschiebungen von Mustern im Eingaberaum als auch Skalierungen dieser Muster sowie moderaten Deformationen umfassen. Einzig Rotationen können nicht kompensiert werden, wobei Publikationen (z. B. [14]) existieren, die diesen Missstand zu beheben versuchen. Das Neocognitron bildet wichtige Eigenschaften des visuellen Wahrnehmungssystems ab und dient so auch als Modell der Erkennungsvorgänge von Mustern im Gehirn. Beispiele hierfür sind die Unterteilung in einfache und komplexe Zellen (S- und C-Zellen) sowie die hierarchische Merkmalsextraktion, die in niederen Schichten primitive Strukturen erkennt, welche in höheren

Schichten zu komplexeren Mustern kombiniert werden, bis eine eindeutige Klassifizierung erreicht wird.

Den positiven Eigenschaften des Neocognitrons stehen auch einige Nachteile gegenüber. Auf der einen Seite sind es die umfassenden Parameter des Systems, die sich meist nur manuell und problemabhängig festlegen lassen. Dazu gehören die Selektivität von S-Zellen, die Verbindungsstruktur zwischen einzelnen Schichten, die Festlegung unveränderlicher Gewichte vor dem Lernvorgang und überhaupt die Anzahl von Schichten, Ebenen und Zellen in diesen. Außerdem kann eine Erkennung, die ähnliche Muster als unterschiedliche Zeichen betrachtet, nur durch einen überwachten Lernvorgang erreicht werden. Doch für diesen ist es notwendig, für jede Ebene des Netzes zu bestimmen, welches Muster durch diese anzulernen ist, ob eine Ebene mehrere Muster lernen soll und ob zwei S-Ebenen durch eine C-Ebene zusammen zu fassen sind. Die Erstellung des Trainingssatzes stellt eine mühevolle und zeitraubende Aufgabe dar, die hochgradig für die Güte der Erkennung verantwortlich ist. Nach Meinung des Autors sind es diese negativen Eigenschaften, welche die problemabhängige Entwicklung eines Neocognitron-Netzwerks zu einer mühevollen Aufgabe machen und dazu führen, dass es in der Literatur so gut wie keine umfangreiche Darstellung gibt, die auch eine Analyse der Eigenschaften des Netzwerks einbezieht.

Mit dem Neocognitron wird einmal mehr unter Beweis gestellt, dass von der Natur inspirierte Systeme zu außerordentlichen Leistungen fähig sind. Verbesserungen, die von Fukushima in jüngerer Zeit eingeführt wurden (siehe [4]), haben das Modell des visuellen Wahrnehmungssystems abermals ausgebaut und zu einer weiteren Verbesserung der Erkennungsleistung durch das Neocognitron geführt. Für die Zukunft sind neueste Entwicklungen interessant, die das Neocognitron mit Rückkopplungen versehen und so in die Lage versetzen, teilweise verdeckte Muster zu rekonstruieren und nachfolgend zu erkennen (siehe [5]). Es bleibt abzuwarten, dass weitere Ergebnisse aus der Gehirnforschung in Modelle neuronaler Netze mit einfließen und durch Simulation dieser eine fruchtbare Wechselwirkung entsteht, die zu einem besseren Verständnis von Wahrnehmungen allgemein und zu umfassenden computergestützten Systemen führt, die in der Lage sind, alle vom Gehirn bekannten Eigenschaften nachzubilden.

A Glossar

Cognitron Vorgänger des Neocognitrons, das 1975 von Prof. Fukushima entwickelt wurde. Dieses KNN war nicht in der Lage, Verschiebungen, Skalierungen und Deformationen zu tolerieren.

Ebene Rechteckiges Feld von Zellen. Eine Schicht des Neocognitrons besteht aus mehreren Ebenen, alle Zellen einer Ebene erkennen dabei dasselbe Muster an unterschiedlichen Positionen im Eingaberaum.

Exzitatorische Verbindung Verbindung, die anregend auf die nachfolgende Zelle einwirkt. Verbindungen von C- zu S-Zellen, von C- zu V-Zellen und von S- zu C-Zellen sind exzitatorisch.

Ganglion-Zelle Ganglienzellen befinden sich in der Retina des Auges in der dritten Schicht und bilden die letzte Stufe vor der Weiterleitung der visuellen Information Richtung Gehirn. Jede Ganglionzelle erhält ihre Eingaben von einem bestimmten Bereich der Retina. Es wird zwischen On-Center und Off-Center Zellen unterschieden. On-Center-Zellen extrahieren positive lokale Kontraste, während Off-Center-Zellen auf negative lokale Kontraste besonders stark reagieren.

Gewicht Jede Verbindung besitzt ein Gewicht, welches multipliziert mit der Ausgabe der Vorgängerzelle als Summand in die Eingabe der Nachfolgerzelle einfließt. Es gibt feste und variable Gewichte, wobei variable Gewichtungen durch den Lernprozess festgelegt werden und das eigentliche Wissen des Netzes speichern.

Graphem Graphische Symbole mit individueller Bedeutung.

Großmutter-Zelle Dies ist eine hypothetische Zelle, welche die Großmutter einer Person oder ein beliebiges komplexeres Objekt repräsentiert. Die Hypothese zielt darauf hinaus, dass die Großmutter-Zelle aktiviert wird, sobald eine Person an das Objekt „Großmutter“ denkt, dass also ein bestimmtes Neuron im Gehirn für eine Entität steht.

Inhibitorische Verbindung Verbindung, die hemmend auf die nachfolgende Zelle einwirkt. Nur die Verbindungen von V- zu S-Zellen sind inhibitorisch und helfen den S-Zellen so, zwischen relevanten und irrelevanten Mustern zu unterscheiden.

KNN → *Künstliches Neuronales Netz.*

Konfiguration Gebiet einer Schicht, dessen Zellen genau eine Zelle jeder Ebene der Nachfolgerschicht speisen. Meist quadratisch oder rechteckig ausgedehnt, es sind jedoch beliebige Formen möglich.

Künstliches Neuronales Netz Abstraktion von natürlichen neuronalen Netzen in Computermodellen. KNNs bestehen aus künstlichen Neuronen, die miteinander verbunden sind und werden vor allem in Gebieten eingesetzt, wo wenig systematisches Wissen über die Problemdomäne existiert.

Kunihiko Fukushima Japanischer Forscher und Professor, Mitbegründer des Cognitrons und Neocognitrons.

Neocognitron Hierarchisches KNN mit besonderen Fähigkeiten zur Mustererkennung, z. B. Toleranz gegenüber Verschiebungen, Skalierungen und Deformationen.

Neuron Im Kontext dieser Arbeit → *Zelle*.

Räumliche Konfiguration → *Konfiguration*.

Retina Synonym für die Netzhaut des Auges. Schicht von Nervengewebe an der hinteren Augen-Innenseite, in der Licht in Nervenimpulse umgewandelt wird, die zum Gehirn weiter geleitet werden.

Rezeptives Feld Eingabeschicht, in welcher die Ausgabe jede Zelle beispielsweise dem Grauwert des ursprünglichen Bildes an dieser Stelle entspricht.

Schicht Besteht aus mehreren Ebenen. Unterscheidung zwischen S- und C-Schichten. Eine Schicht ist zur Erkennung mehrerer unterschiedlicher Muster verantwortlich, die jedoch von gleicher Ordnung (ähnlicher Komplexität) sind.

Selektivität Faktor, der die Hemmung von S-Zellen durch ihre V-Zellen bestimmt. Je größer die Selektivität ist, desto höher ist die Trennschärfe, desto deformationsintoleranter ist eine S-Zelle jedoch auch.

Stufe Gröbster Bestandteil des Neocognitrons, welches aus mehreren Stufen aufgebaut ist. Eine Stufe besteht aus jeweils einer S-Schicht, gefolgt von einer C-Schicht.

Trennschärfe Eigenschaft des Neocognitrons zwischen unterschiedlichen Mustern zu unterscheiden. Es besteht dabei eine Dualität zwischen Trennschärfe und Deformationstolerierung, die u.a. durch die Selektivität bestimmt wird.

Verbindung Verbindet zwei *Zellen* miteinander. Jede Verbindung besitzt ein positives *Gewicht*. Dieses kann fest vorgegeben oder variabel sein.

Visueller Kortex Auch „Sehrinde“ genannt. Teil der Großhirnrinde, der zum visuellen System zählt. Ermöglicht die visuelle Wahrnehmung im Gehirn.

weight sharing Prinzip des Neocognitrons, bei dem sich alle S-Zellen einer Ebene dieselben Eingangs-Gewichte teilen, sodass diese nur einmal pro Ebene abgespeichert werden müssen. S-Zellen einer Ebene reagieren damit auf dasselbe Muster, nur an unterschiedlichen Positionen im Eingaberaum.

Zelle Im Kontext dieser Arbeit ein künstliches Neuron im Neocognitron-Netzwerk. Eine Zelle ist mit m Zellen verbunden, von denen sie ihre Eingabe erhält und mit n Zellen, denen sie ihre Ausgabe zur Verfügung stellt.

Literatur

[1] Fukushima, Kunihiko: Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. In: *Biological Cybernetics* 36 (1980), S. 193–202

[2] Fukushima, Kunihiko: Neocognition: A Hierarchical Neural Network Capable of Visual Pattern Recognition. In: *Neural Networks* 1 (1988)

[3] Fukushima, Kunihiko: Analysis of the Process of Visual Pattern Recognition by the Neocognitron. In: *Neural Networks* 2 (1989)

[4] Fukushima, Kunihiko: Neocognitron for handwritten digit recognition. In: *Neurocomputing* 51 (2003), S. 161–180

[5] Fukushima, Kunihiko: Restoring partly occluded patterns: a neural network model. In: *Neural Networks* 18(1) (2005), S. 33–43

[6] Fukushima, Kunihiko: Use of non-uniform spatial blur for image comparison: symmetry axis extraction. In: *Neural Networks* 18(1) (2005), S. 23–32

[7] Fukushima, Kunihiko: Neocognition: A Neural Network Model for a Mechanism of Visual Pattern Recognition. In: *IEEE Transactions on Systems, Man, and Cybernetics* 13(5) (September/October 1983), S. 826–834

[8] Fukushima, Kunihiko ; Wake, Nobuaki: Handwritten alphanumeric character recognition by the Neocognitron. In: *IEEE Transaction on Neural Networks* 2(3) (1991), S. 355–365

[9] Hecht-Nielsen, Robert: *Neurocomputing*. Addison-Wesley Publishing Company, Inc., 1990. – ISBN 0–201–09355–3

[10] Kuhlins, Stefan: Neocognitron (Seminararbeit). (1996). `http://www.informatik.uni-mannheim.de/pi4.data/content/courses/1995-ws/seminarWS9596/StephanKuhlins.ps.gz`

[11] Lovell, David R. ; Downs, Thomas ; Tsoi, Ah C.: An Evaluation of The Neocognitron. In: *IEEE Transactions on Neural Networks* 8(5) (1997), 1090-1104. `http://citeseer.ist.psu.edu/lovell97evaluation.html`

[12] Lovell, David R. ; Simon, David ; Tsoi, Ah C.: Improving the Performance of the Neocognitron. In: *Fourth Australian Conference on Neural Networks* (1993), 22-25. `http://citeseer.ist.psu.edu/78661.html`

[13] Patterson, Dan: *Künstliche neuronale Netze: das Lehrbuch*. Prentice Hall Verlag GmbH, 1996. – ISBN 3–8272–9531–9

[14] Satoh, Shunji ; Kuroiwa, Jousuke ; Aso, Hirotomo ; Miyake, Shogo: Recognition of Rotated Patterns Using Neocognitron. In: *Proc. Int. Conf. Neural Information Processing* 1 (1997), 112-116. `http://citeseer.ist.psu.edu/satoh97recognition.html`

[15] Velinsky, Tomas: *Neocognitron*. `http://cs.felk.cvut.cz/~neurony/neocog/en/`, letzter Zugriff: 24.03.2007

[16] Zell, Andreas: *Simulation Neuronaler Netze*. Addison-Wesley (Deutschland) GmbH, 1994. – ISBN 3–89319–554–8