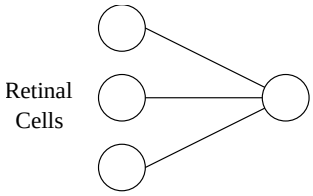
Modelling the Development of the Cat Lateral Geniculate Nucleus with Hebbian Learning

Stephen Eglen

School of Cognitive and Computing Sciences



Geniculate Normalisation

$$w'_{ij} = w_{ij} - \frac{1}{nR}\left(tL - \sum_{k=1}^{nL} w_{kj}\right)$$

Subtractive normalisation is actually an iterative procedure, since there is an additional constraint that

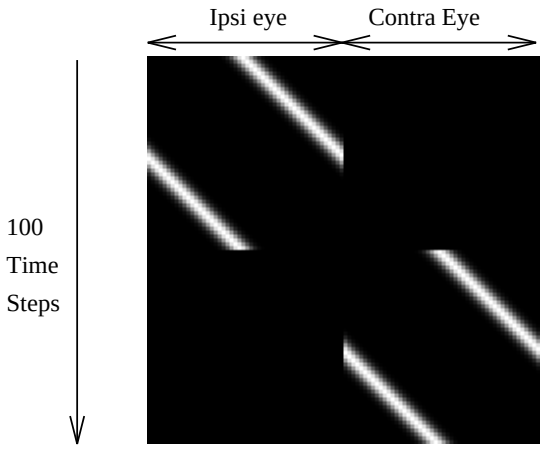


Figure 5: Retinal inputs shown over one epoch (100 iterations). In this example, only one eye is active at a time.

normalised either after every iteration (presentation of one input vector) or after every epoch (presentation of all input vectors). Growth learning takes place at random, typically so that it happens once or twice an epoch.

3 Results

This section describes the experiments performed with the network. Figure 6 explains how to interpret the weight diagrams shown in this paper. Hinton diagrams are also shown, but these have only been used on a subset of the weight matrix for reasons of space. Ocular dominance plots are also used – these plots colour each LGN cell according to the nature of its inputs. LGN cells receiving at least 80% of its weights from the ipsilateral (contralateral) eye are coloured black (white). All other cells are coloured grey to indicate they receive both ipsilateral and contralateral input.

Five sets of experiments are described here, listed below. Tests 1 to 3 replicate the results described in (Keesing et al., 1992). The remaining tests examine certain aspects of this model.

1. Simultaneous arrival of all retinal afferents.
2. Biased arrival times of retinal afferents.
3. Chemical Gradients used to Bias Topology.
4. Importance of growth term.
5. Importance of normalisation.

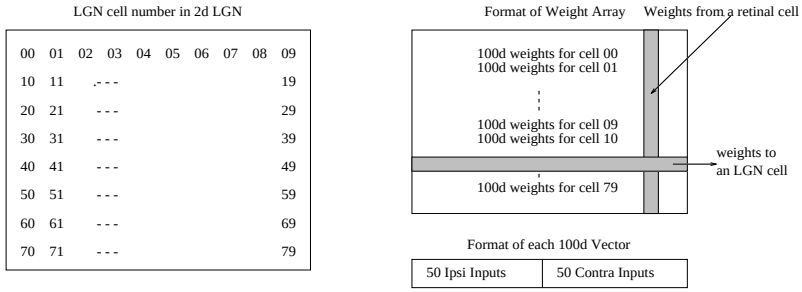


Figure 6: Interpreting the weight diagrams. Each LGN cell is numbered as shown, from 0 to 79. The right hand figure shows how the weights are displayed: row n of the weights array corresponds to the 100 dimensional vector for LGN cell n. Each weight vector can be subdivided into the 50 weights for the ipsilateral eye and 50 weights for the contralateral eye. Each element of the weight array is represented by a pixel: the larger the weight, the brighter the pixel.

3.1 Test 1: Simultaneous Arrival of All Retinal Afferents

3.2 Test 2: Biased Arrival Times of Retinal Afferents

The previous test showed that given completely random initial weights, global topography and layered ocular segregation do not develop. It is known however that contralateral eye inputs innervate the LGN several days earlier than ipsilateral eye inputs (Shatz, 1994, p535). This means that the contralateral inputs have an advantage in innervating the upper half of the LGN, namely layer A. (This earlier arrival time for contralateral inputs explains why the same pattern of ocular segregation always arises in the cat LGN.)

To model this difference in arrival times, we can bias the initial weights so that only the contralateral inputs can initially innervate the upper half of the LGN model. To do this, we initialise the weights to random values as before, and then set the weights from the ipsilateral eye to the teshotlf og

3.3 Test 3: Chemical Gradients used to Bias Topology

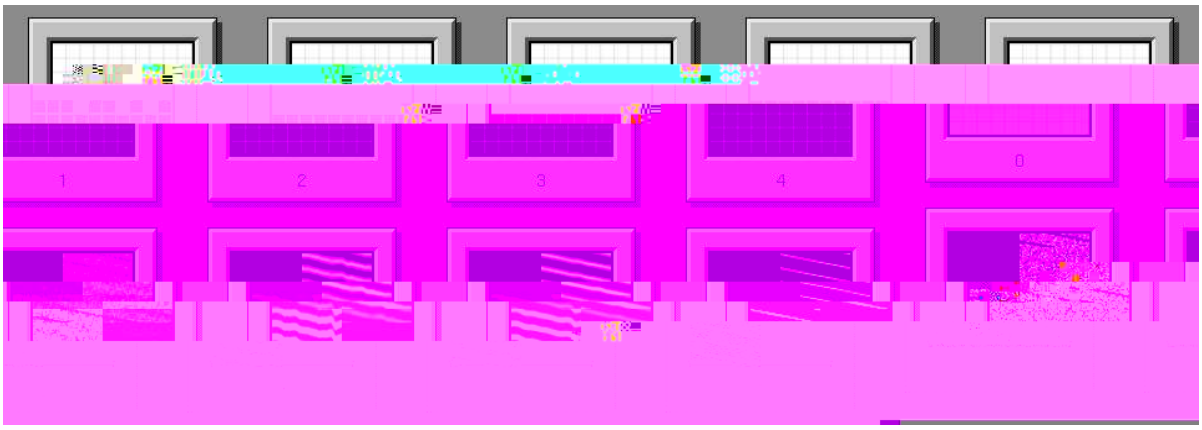
So far, the model can generate local topographic mappings, but there is no global mapping from the retina to each row of the LGN. As argued by Willshaw and Von der Malsburg (1979), some initial specification of the orientation information must be required (for example, which end of the retina connects to which end of the LGN), as there is nothing in the model to predispose it to one particular orientation.

An extra mechanism to provide this global mapping is likely to exist in the retinogeniculate pathway: it is widely thought that chemical gradients are responsible for guiding the retinal ganglion cell axons to the LGN and then projecting in a coarse grained topographic manner (Goodman & Shatz, 1993, p92).

To incorporate this mechanism into the model, the initial weights of the network are biased so that the weights at an intermediate distance away from the main diagonal of a row of weights are set to zero, or to small random values, as shown in Figure 10.

It is important to note when biasing the weights for topography, there is no need to bias all of the weights: it is sufficient to bias just one row in each of the contralateral / ipsilateral parts of the weight array. Once there is a bias in one row for each eye, the growth rule ensures that this bias is transferred into the other rows of the LGN.

Figure 11, picture 0 shows an initial set of weights which are biased for both topography and ocular segregation: the contralateral eye innervates all eight rows of the LGN, whereas the ipsilateral eye innervates only the bottom four rows. As can be seen from this sequence of pictures, the network transfers the bias


1
2
3
4
0

3.4 Test 4: Importance of Growth Term

This test investigates the importance of the growth term in the learning rule: what happens if there is no growth learning? Figure 13 shows the results of a typical experiment when there is no growth learning in the model. As can be seen, although the ocular segregation is formed, the topography completely breaks down on both a global and a local scale. This is not surprising, given the nature of the growth term. It seems to have two responsibilities:

- Make sure that neighbouring LGN cells become responsive to s ar

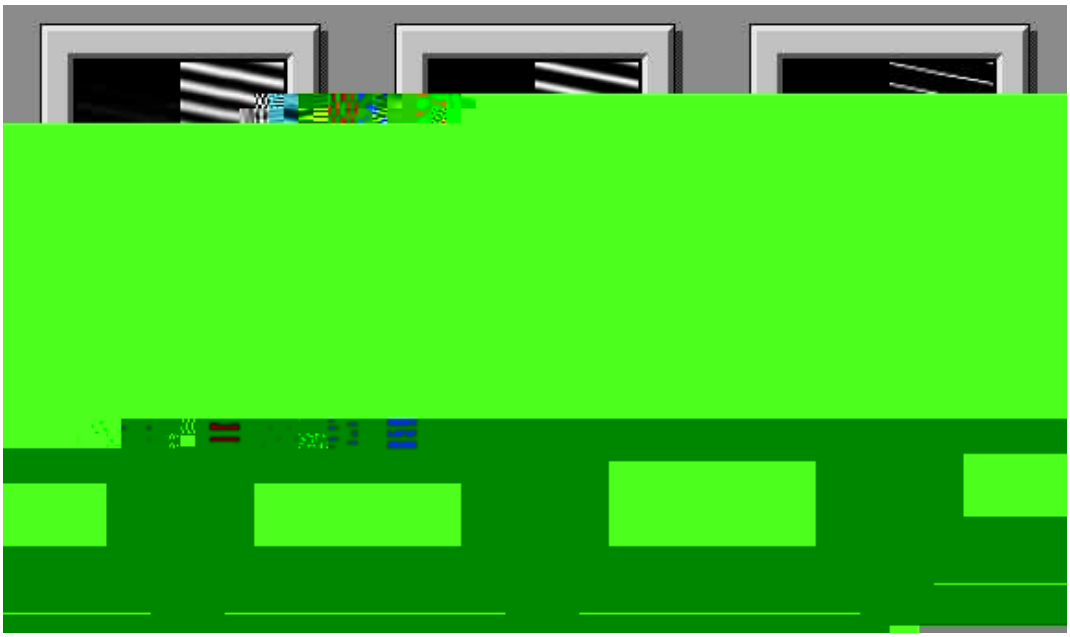


Figure 14: Effect of the radius parameter of the neighbourhood. (0) Weights after 500 epochs of training with $radius$ equal to 2, (1), $radius$ is then decreased to 1 for another 500 epochs, and (2) finally $radius$ is decreased to 0 and the network trained for another 500 epochs. The smaller the final value of $radius$, the more refined the topography is. As shown by the corresponding ocular dominance plots, the binocular region in the middle of the LGN disappears only when rad = 0.

3.5 Test 5: Importance of Normalisation

In the original model, two methods of normalisation were imposed: divisive retinal normalisation and subtractive geniculate normalisation. An important question is whether both these normalisation schemes are necessary for proper development. The normalisation methods have very different properties, as shown in (Miller & Mackay, 1994; Goodhill & Barrow, 1994). Does the network rely on particular features of each normalisation technique in order to develop, or can it use any normalisation technique? Furthermore, are the two forms of normalisation (geniculate and retinal) normalisation required? If normalisation is being used here just to keep the weights bounded, then only one normalisation scheme is sufficient.

These questions have been investigated by systematically varying the form of normalisation used at each site. Normalisation can either be subtractive or divisive at each site, or alternatively there can be no normalisation. This gives $3 \times$

divisive. If subtractive retinal normalisation is used, then the topography is lost, although every retinal cell connects to the LGN, and layered ocular segregation is often

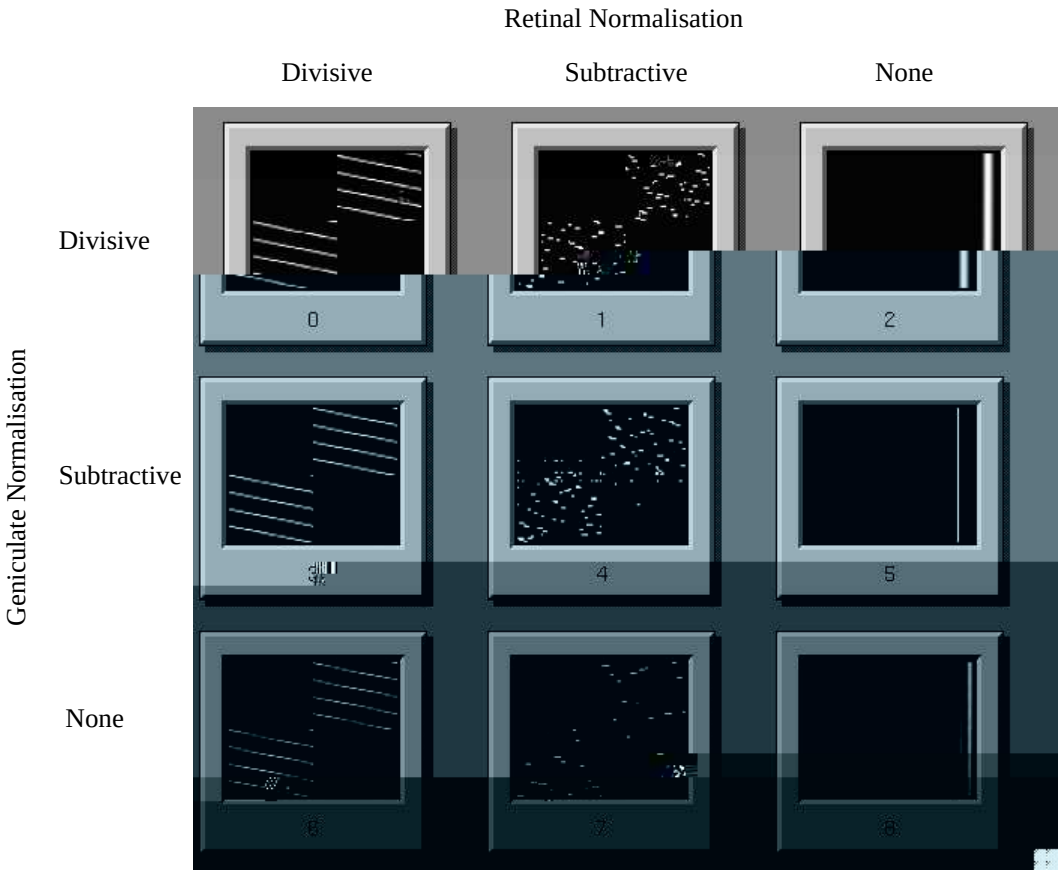


Figure 15: Final weight diagrams for the normalisation tests listed in Table 3. Note: In picture 8, the weights are free to grow without bounds and eventually cause very large weights, since there is no upper bound on individual weight strengths. (Weights are still co

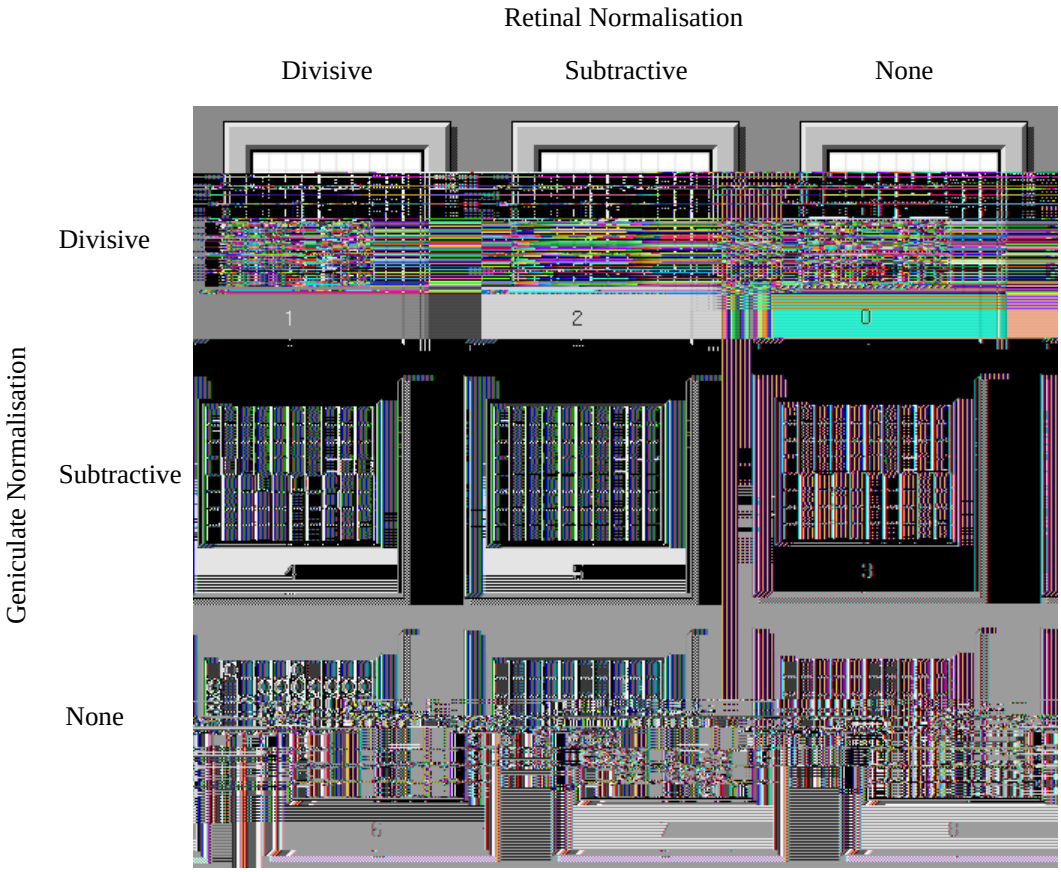
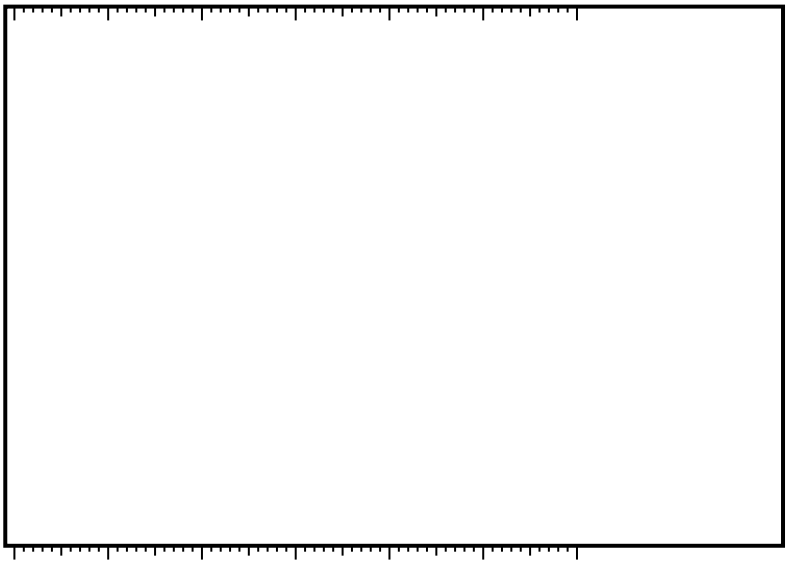


Figure 16: Final ocular dominance plots for the normalisation tests listed in Table 3. In picture 7, the cells



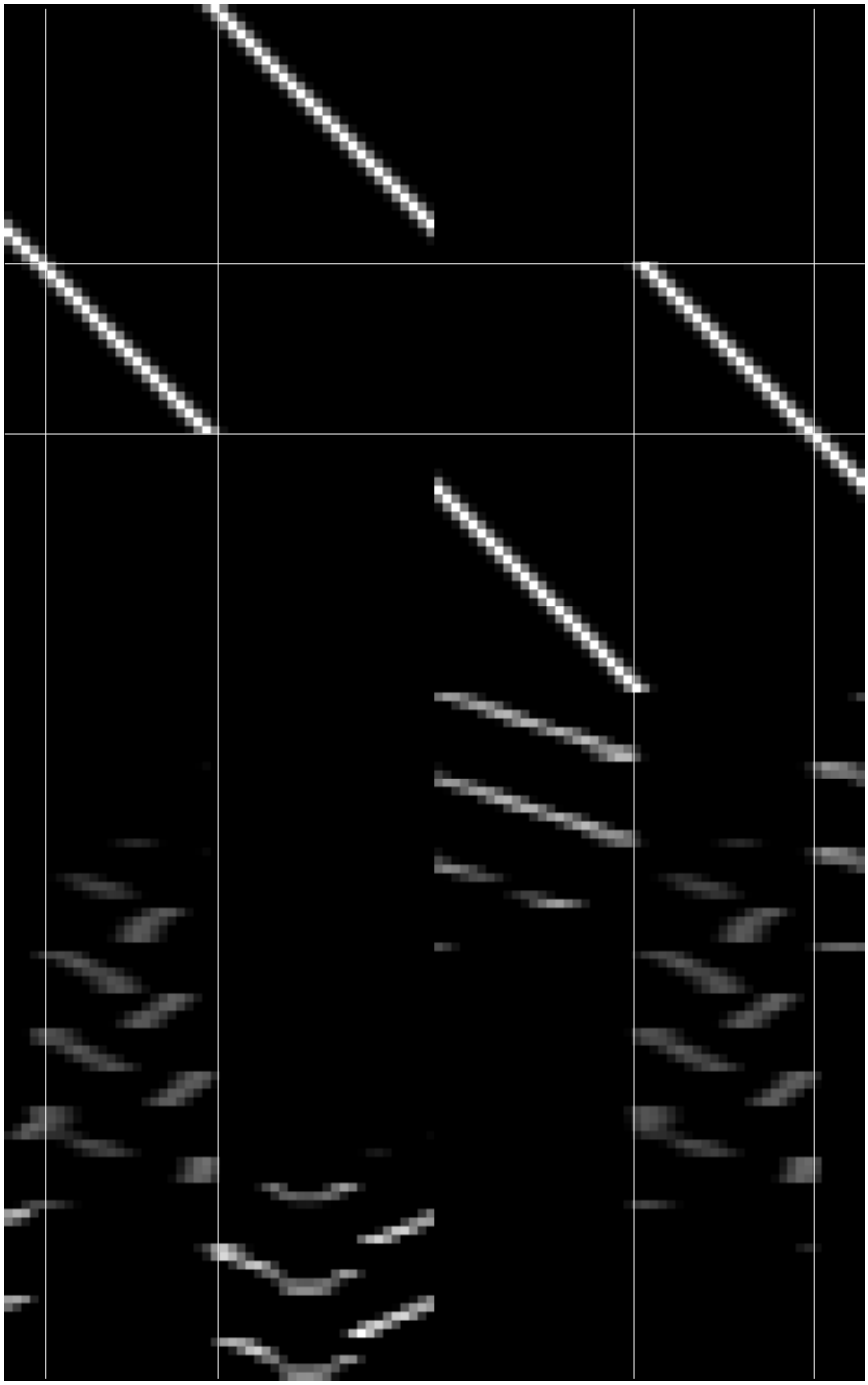
3.5.1 Is geniculate normalisation necessary?

Random Overlapping Waves

To investigate whether this model could develop normally using inputs that do have a period of overlap when both eyes s bt12.86(p)-2 m12.8625(l)-9.42696(o,32.933344)-2.9476(pn2.8625(l)-9.4269)-2(s)7.949o nr-2.42s17p nhag


2
3

Area of Overlap



Weight Diagram

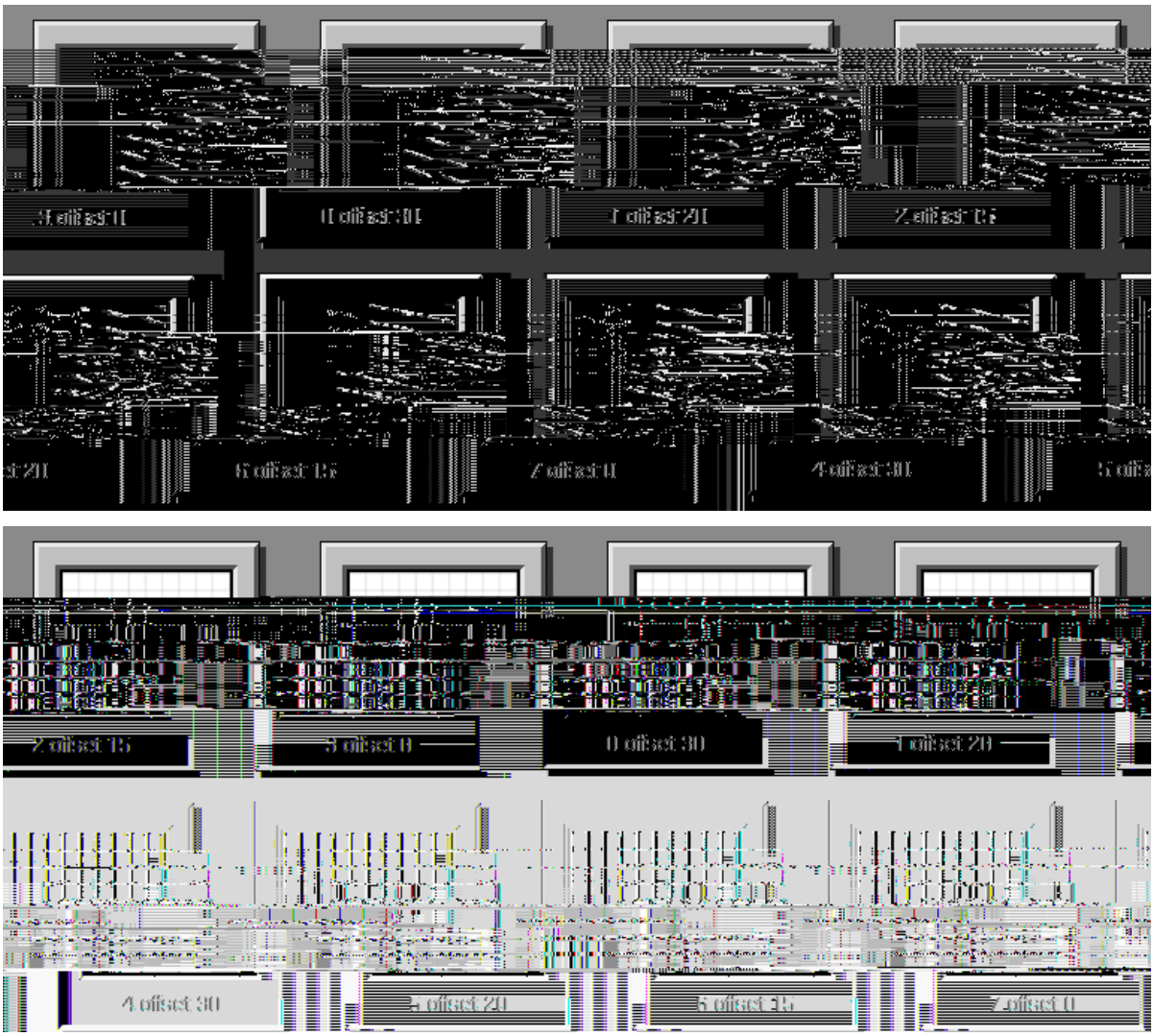


Figure 24: Final weight patterns using02a7(2)-2.1u((F)5.10211(i)-9.42i0.(n)-2.9945 (F)5.10211(i)-9.4(e)121.8535(r)-2.39694h

Fixed Overlapping Waves

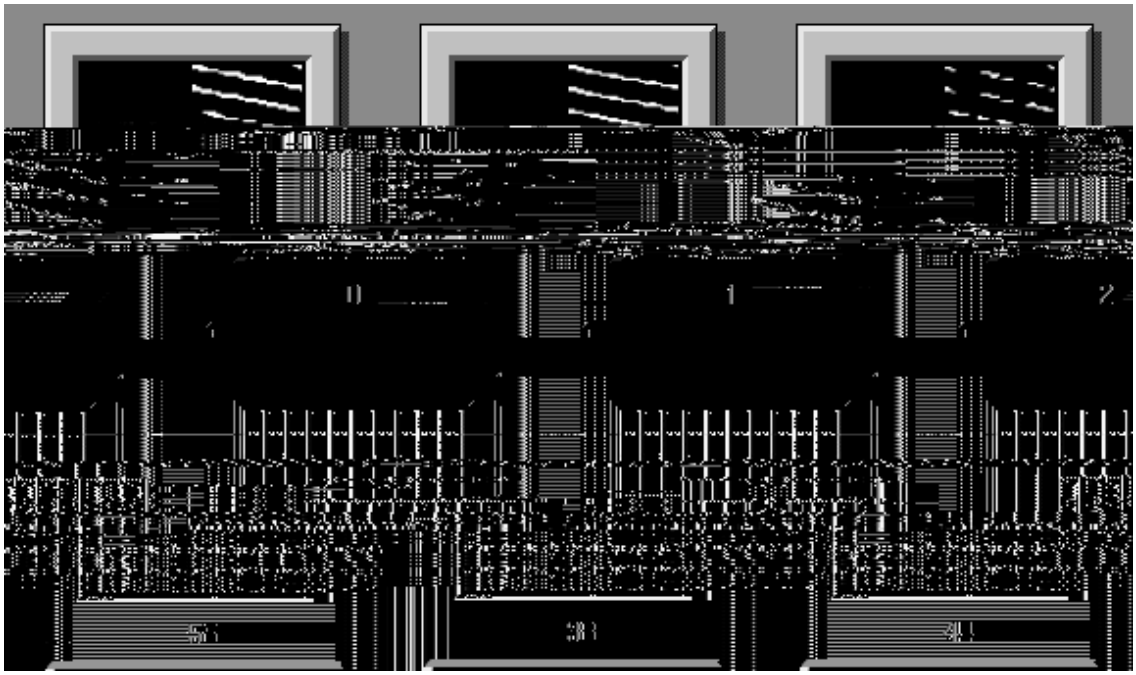


Figure 26: Effect of learning with subtractive retinal normalisation constrained also such that each weight is bounded in the range $[0, maxWt]$. No geniculate normalisation, (0) rs = 1.0, $maxWt$ = 0.20 (1) rs = 0.2, $maxWt$ = 0.2, (2) rs = 0.2, $maxWt$ = 0.50. Pictures 3,4 and 5 show the corresponding ocular dominance plots. Non overlapping retinal inputs were used for these experiments, but experiments with the random overlapping inputs produced similar results.

Satisfying both forms of normalisation.

When both retinal and geniculate normalisation is being used in the model, there is the problem of how to satisfy both forms of normalisation at the same time. If the normalisations are applied sequentially, there is no guarantee that imposing the second normalisation will not destroy the constraints imposed by the first normalisation. Some tests were therefore performed to see how model performance is affected by the interactions between the two forms of normalisation.

In the previous experiments, at the end of every epoch, retinal normalisation was applied first followed by geniculate normalisation. In this set of experiments, the probability of geniculate normalisation occurring first (pN) was allowed to vary between 0 (divisive retinal normalisation is always applied first for this epoch) and 1 (subtractive geniculate normalisation is always applied first for this epoch). The results are shown in Figure 27: it is clear that the normal topography develops only if the divisive retinal normalisation is applied before the subtractive geniculate normalisation. (In contrast, both Miller et al. (1989) and Goodhill (1992) apply the output cell normalisation before the input cell normalisation.)

Why does applying subtractive geniculate normalisation first produce strange results?

One possibility why applying subtractive geniculate norma

Secondly, Miller and Mackay (1994) reported that the form of

A Parameters Used.

Table 5 shows the parameters of the model, along with typical values.

Parameter	Use/Meaning	Value
α	Average Input Value, used in learning rule	0.1
β	Average Output Value, used in learning rule	0.0125
γ	Constant for growth term	0.1
$radius$	radius of neighbourhood	[0,1,2]
nL	Number of LGN cells	80
nR	Number of Retinal cells	100
tR	Target Sum for weights from each Retinal Cell	1.0
tL	Target sum for weights to each LGN cell	1.25

Table 5: Parameters and typical values used in the simulation.

There is a relationship between the target sums tR and tL:

$$\text{average weight value} = \frac{tR}{nl} = \underline{\quad tL \quad}$$