What is the optimal architecture for visual information routing?

Philipp Wolfrum and Christoph von der Malsburg

December 12, 2006

Abstract

Analyzing the design of networks for visual information routing is an underconstrained problem due to insufficient anatomical and physiological data. We propose here optimality criteria for the design of routing networks. For a very general architecture we derive the number of routing layers and the fanout that minimize the required neural circuitry. The optimal fanout l is independent of network size, while the number k of layers scales logarithmically (with a prefactor < 1) with the number n of visual resolution units to be routed independently. The results are found to agree with data of the primate visual system.

1 Introduction

An impressive capability of biological vision systems is invariant object recognition. The same object seen at different position, distance, or under rotation leads to entirely different retinal images which have to be perceived as the same object. Only one of these variances, translation, can be compensated by movements of the eye. The approximate log-polar transform which takes place in the mapping from retina to cortex (Schwartz, 1977), on the other hand, replaces some kinds of transformations (scale, rotation) by others (translation on the cortex) and therefore does not fully explain invariant recognition, either. Invariant recognition, and how the visual system performs it, remains a topic far from being understood.

Invariance does *not* mean insensitivity to the spatial arrangement of visual information. While object recognition in our brain is invariant with respect to the above-mentioned transformations, it is very sensitive to small differences in the retinal activity pattern arising from, say, seeing both of your twin sisters shortly after each other. We believe that the only way a brain can solve these two competing problems realistically is to have a general, object-independent mechanism that compensates variance transformations without distorting the image to convey a normalized version of it to higher brain areas for recognition. Such a mechanism requires a routing network providing physical connections between all locations in the visual input region (V1) to all points in the target area (like IT). Additionally, neural machinery is required to control these connections.

The necessity for dynamic information routing was appreciated early on in vision research (Pitts and McCulloch, 1947), and several architectures have been proposed, like Shifter Circuits (Olshausen, Anderson and van Essen, 1993) or the SCAN model (Postma, van den Herik and Hudson, 1997). What has been missing so far is a discussion of efficiency in terms of required neural resources. Different routing architectures require different numbers of intermediate feature-representing nodes (we will refer to them simply as *nodes*) and node-to-node connections (*links* from now on; if we mean both links and nodes, we will use the term *units*). While we will not discuss in this paper *how* connections in a routing circuit are controlled (for this, refer to (Olshausen et al., 1993) and to (Lücke, 2005)), we assume that the maintenance of a link and its control by a neural control unit have the same cost as feature nodes (for a deviation from this assumption, see Sect. 2.1). It is likely that cortical architectures have evolved which minimize this cost for the organism. In this paper, we therefore derive and analyze the routing network structure that minimizes the sum of all required units, both nodes and links.

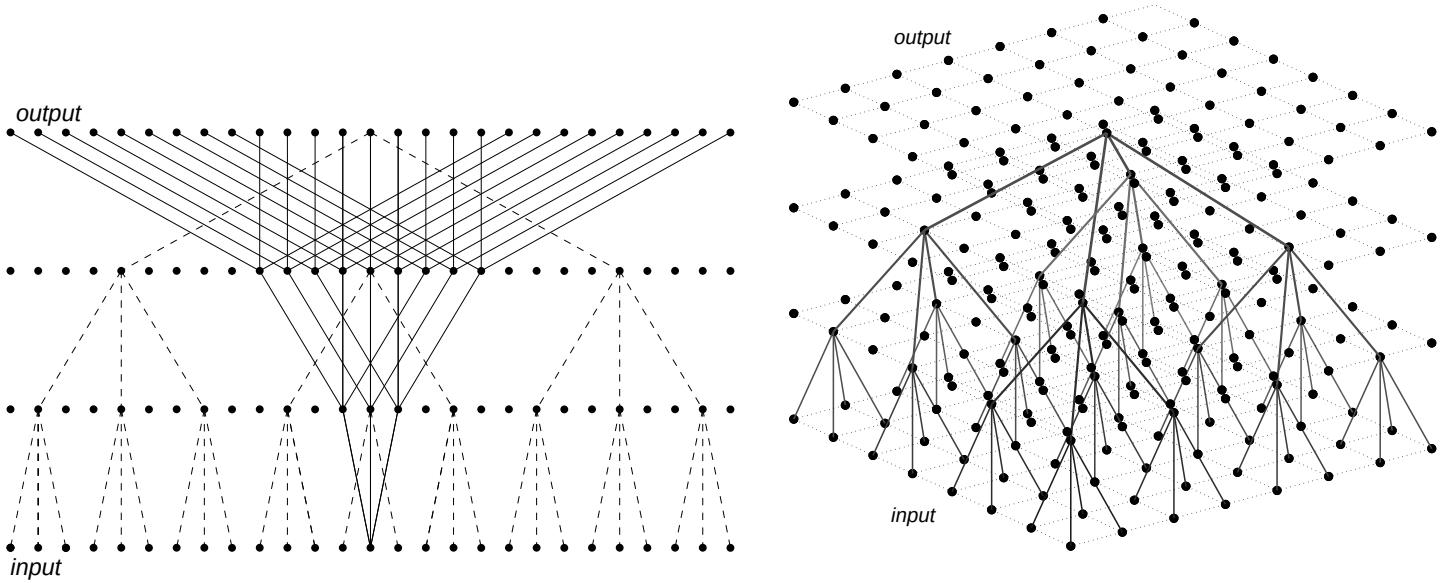
In our analysis we will focus on two situations. In Sect. 2 we discuss routing between two cortical regions of identical size. This corresponds to perception of an already coarsely segmented object. In Sect. 3 we consider the architecture that must be present in real biological vision systems: Routing from a large input domain to a much smaller output domain, the first corresponding to the whole visual field, the second to a small higher-level target area engaged in object recognition. After analysis of these two cases we interpret our results quantitatively for the case of the primate brain (Sect. 4) and discuss some implications and predictions arising from them (Sect. 5).

2 Routing between two regions of the same size

Let us define a routing architecture with as few assumptions as possible:

- Input and output stages both consist of n image points. Each image point is represented by one feature unit (the extension of this to more than one feature per image point will be discussed below).
- The routing between input and output is established via $k-1$ intermediate layers of n feature units each.
- Nodes of adjacent feature layers can be connected. For every such connection there exists one dynamic, neural unit that controls information flow in both directions. These units resemble the "control units" of (Olshausen et al., 1993). We assume here that one link (including its control unit) imposes the same "maintenance cost" as one feature node. If these costs are not identical, this can be accounted for with the parameter α introduced in Sect. 2.1.

Under these assumptions, what is the minimal architecture providing for each input node one separate pathway to every output node? For $k = 1$, the situation is clear: Without any intermediate layer, every input node must be connected to all n output nodes. With intermediate layers, however, we can make use of a combinatorial code to achieve full connectivity, similar to "butterfly" computations used in the fast Fourier transform (**?**): We assume that each input unit is only connected to l nodes of the adjacent intermediate layer (see solid lines in Fig. 1(a)). Each of these l nodes has in turn connections to l nodes of the following layer, and so on, until the output stage is reached. This method yields for every input node l^k pathways to the output stage, which are unique and lead all to different output nodes *if* we make sure that no two separate pathways merge again on the way to the output stage. An anatomically plausible way to meet this functional requirement is to let the spacing between target points increase exactly by the factor l from one link layer to the next, as shown in Fig. 1(a). The two-dimensional case is analogous, except that here the groups of nodes projecting to the same target are two-dimensional patches of l units with adequate spacing in between (see Fig. 1(b)).



(a) (b)

Figure 1: Architectures for routing networks. (a) The one dimensional case, with $n = 27$ and $k = 3$, thus $l = n^{\frac{1}{k}} = 3$. All feature nodes are shown (dots), but only selected links (lines), the others being shifted versions (with circular boundary conditions) of the shown links. All connections from one input node to the whole output stage are shown as solid lines, the connectivity between one output node and the whole input as dashed lines. (b) The two dimensional case, with $n = 64$, $k = 3$, and $l = 4$. Only the downward connections from a single output node are shown.

The connectivity described here agrees with the general anatomical finding that the spread of neuronal connections increases along the visual hierarchy. Perkel et al. (Perkel, Bullier and Kennedy, 1986), for example, found a higher divergence of projections between V1 and V4 than between V1 and V2 or V3, respectively. In (Tanigawa, Wang and Fujita, 2005), a four times larger spread of horizontal axons in inferotemporal cortex than in V1 was reported. Note, however, that the specific connectivity of the routing network is irrelevant for the results derived in the following. The only requirement is that the pathways of every input node be unique and lead to different output nodes.

In order to reach the whole output stage with these pathways, their number must equal the number of output nodes:

$$n = l^k$$

From this we get the necessary neuronal fan-out at each stage as

$$l = n^{\frac{1}{k}}. \tag{2.1}$$

Let us now calculate how many nodes are needed to realize the routing architecture. Having $k-1$ intermediate layers means that a total of $(k+1)n$ feature nodes is required. All of these nodes, except those of the output layer, have l links to the next stage, resulting in a total of $knl = kn^{\frac{k+1}{k}}$ links. So the total number of units as a function of k and n is

$$N(k,n) = (k+1)n + kn^{\frac{k+1}{k}}. \tag{2.2}$$

As we can see in Fig. 2, this number changes drastically with the number of intermediate layers being used. A direct all-to-all connectivity without any intermediate layers ($k = 1$) is most expensive because the number of required links scales quadratically with n in this case. For a very large number of intermediate layers, on the other hand, the decrease in the required fan-out l is outweighed by the linear increase in nodes caused by additional layers. As we can see, there is a unique value k_{opt} for which the number of required units attains a minimum. To determine k_{opt}, we calculate the derivative of N with respect to k and set it to zero:

$$\frac{\partial N}{\partial k} = n + n^{\frac{k+1}{k}} - k\ \ln n \frac{1}{k^2} n^{\frac{k+1}{k}} = n\left(1 + n^{\frac{1}{k}} - n^{\frac{1}{k}} \frac{\ln n}{k}\right) \stackrel{!}{=} 0$$

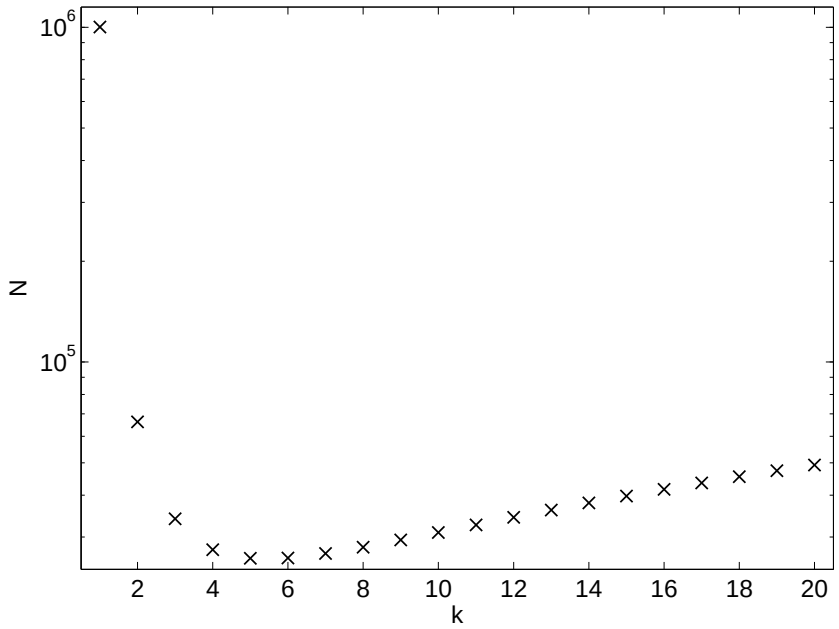


Figure 2: The number of required units for a routing architecture between two layers depends strongly on the number $k-1$ of intermediate layers being used. The values shown here are for input and output stages of $n = 1000$ image points each.

This is satisfied for

$$\frac{\ln n}{k} = n^{-\frac{1}{k}} + 1. \tag{2.3}$$

With the ansatz

$$k_{opt} = c \ \ln n \tag{2.4}$$

(2.3) becomes independent of n:

$$\frac{1}{c} = e^{-\frac{1}{c}} + 1. \tag{2.5}$$

Solving this numerically we obtain

$$k_{\text{opt}} \approx 0.7822 \ \ln n. \tag{2.6}$$

The fact that k_{opt} scales logarithmically with n is not surprising by itself. Such a scaling behavior lies at the heart of many techniques that have to permute or operate on a group of nodes simultaneously, like permutation networks or the fast Fourier transform (**?**). Even in random graphs the network diameter (corresponding somewhat to our number of layers k) scales logarithmically with the number of nodes. This general logarithmic scaling behavior is independent of the specific fanout (or degree) at each node. A different fanout only changes the basis of the logarithm, which is equivalent to changing the prefactor in the logarithmic relation. Here, however, minimizing the number of components of the network leads to a *specific* logarithmic scaling, or phrased differently: The prefactor c in k_{opt} is unique. This goes hand in hand with the existence of a unique optimal fanout

$$l_{\text{opt}} = n^{\frac{1}{k_{\text{opt}}}} = e^{\frac{1}{c}}. \tag{2.7}$$

We will discuss this finding further in Sect. 5.

2.1 More than one feature per image point

So far, we have neglected the routing of visual information when there are several feature cells at one image point. Instead of a dense pixel array, visual information in V1 is represented by a pattern of "hypercolumns" of lower density. However, each hypercolumn contains cells responsive to many different local properties of the input, such as wavelet-like features (Gabors) at different orientations and spatial frequencies, different colors or specificity for one eye or the other.

It may not be necessary to route these features independently of each other to higher areas, so one might assume that only one active link is needed to route many feature units in one image location. On the other hand, certain feature types do require individual treatment. For example, for full orientation invariance, units of one orientation specificity of the input would need connections to all orientation specificities of the output domain. The truth probably lies somewhere in between these two extremes, as suggested in (Zhu and von der Malsburg, 2004): Image points are not routed individually, but in small assemblies through collective links called "maplets". For every group of nodes, there exist several such maplets, responsible for routing at different scales and orientations without requiring individual links for all features in all positions.

Since the focus of this paper is not on a specific routing architecture, but on finding the optimal number of layers for a very general architecture, we will merge the above arguments into a single factor $\alpha \geq 1$ representing the number of feature nodes that are controlled by a single link. If necessary, the parameter α can also be used to account for unequal expense assumed for feature units versus link units.

Instead of n independent feature nodes, we now have $n_\alpha = \frac{n}{\alpha}$ groups of nodes, each containing α nodes. With this, the number of units in the routing circuit (2.2) changes to

$$N(k,n) = \alpha(k+1)n_\alpha + k n_\alpha^{\frac{k+1}{k}}. \tag{2.2'}$$

Setting the derivative of (2.2') to zero leads to

$$\frac{\ln n_\alpha}{k} = \alpha n_\alpha^{-\frac{1}{k}} + 1 = 0. \tag{2.3'}$$

The new ansatz

$$k_{opt} = c \ln n_\alpha \tag{2.4'}$$

yields

$$\frac{1}{c} = \alpha e^{-\frac{1}{c}} + 1, \tag{2.5'}$$

which we can solve numerically for explicit values of α. In Fig. 3 we see that c—and with it k_{opt}—only changes by a factor of 2 over a reasonably large range of α.

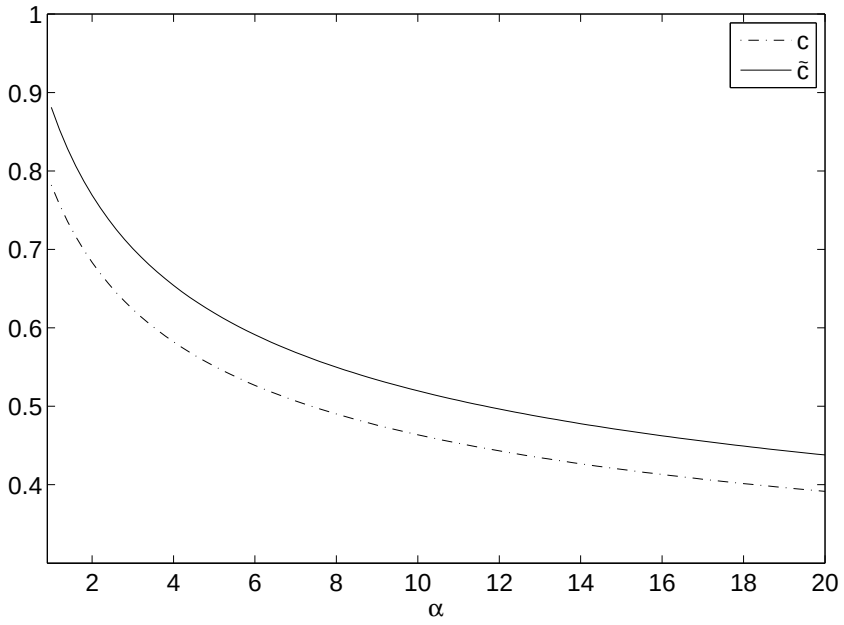


Figure 3: The prefactors c and $\tilde{c}$ define k_{opt} through (2.4') and (3.3) in the cases of routing to an output of the same size or much smaller size, respectively.

Having determined the number of layers k_{opt} that minimizes the required neural circuitry for given n and α, we can calculate the size N_{opt} of this minimal circuitry. Inserting (2.4') into (2.2') yields

$$N_{opt}(n) = n + \left(\alpha + e^{\frac{1}{c}}\right)(cn_\alpha \ln n_\alpha)\,. \tag{2.8}$$

This means that for large n the number of units of the optimal routing architecture between two layers of n_α image points scales with

$$N_{opt}(n_\alpha) \propto n_\alpha \ln n_\alpha, \tag{2.9}$$

as expected from classical network theory. This result holds also for routing of only a single feature ($\alpha = 1$) per point.

3 Routing circuit with different sizes of input and output layer

Let us now discuss routing from the whole visual field to a comparatively small cortical output region. We assume that an attentional mechanism singles out, in the input domain, a region that is to be mapped to the output region.

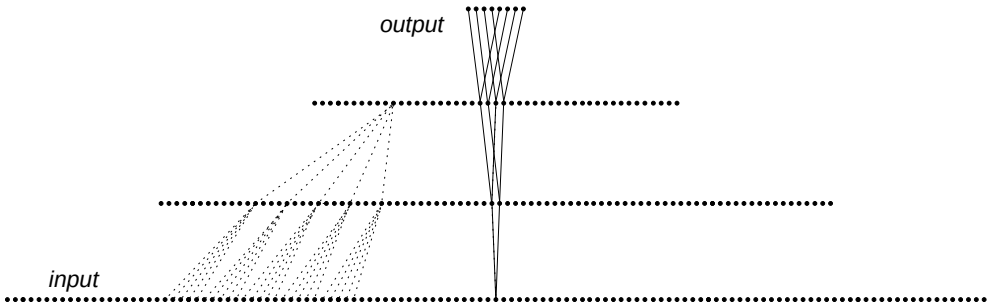


Figure 4: Routing network for an input of $n = 125$ and an output of $\frac{n}{m} = 8$ nodes via $k = 3$ link layers. Consequently, the upward fan-out is $l = 2$. The full lines show the links connecting an input node with the full output stage. Downward connectivity is $l_{down} = n^{\frac{1}{3}} = 5 > l$ (shown exemplarily for a node on the second level of the architecture by dotted lines).

We do not claim here that invariant recognition is perfect over the whole visual field (there are studies showing that this is not the case (Dill and

Fahle, 1998; Cox, Meier, Oertelt and DiCarlo, 2005)), but object recognition *is* possible to some degree even at high retinal eccentricities, although of course impaired by the poor resolution at these parts of the retina. In any case, the basic problem remains the same as before: Dynamic connections must exist between all parts of the visual input region and a target area.

Computationally, the situation is very similar to the one discussed in the previous section and leads to a generalization of the results derived there. We now want to connect an input stage of n units with an output stage that is smaller by the factor m and contains only $\frac{n}{m}$ units. As before, the routing is established via $k-1$ intermediate layers, and groups of α nodes can be routed collectively. With the same argument as in Sect. 2 we see that now each group has to make

$$l = \left(\frac{n_\alpha}{m}\right)^{\frac{1}{k}} \tag{3.1}$$

connections to the next higher layer in order to connect every group of input nodes with every output group. Fig. 4 shows parts of the architecture required for routing from a 125 node input to an 8 node output stage. Note that due to the different input and output sizes downward fan-out now has to be higher than the upward fan-out l.

Differently from before, the size of intermediate layers is not well-defined now. We will assume that the number of nodes changes linearly from the input to the output layer. This is supported by measurements of the average sizes of primary visual areas in humans (Dougherty, Koch, Brewer, Fischer, Modersitzki and Wandell, 2003). Note, however, that the same paper reports variance of V1 sizes of more than 100% between different individuals. In general there seems to be little undisputed data on this question in the literature. Given this uncertainty in the anatomical data, the simplest possible assumption is probably best for this kind of general discussion.

With a linear decrease in size, the number of feature units in layer κ ($\kappa = 0$ for the input and $\kappa = k$ for the output layer) is

$$f_\kappa = n - \frac{\kappa}{k}\left(n - \frac{n}{m}\right).$$

The number F of the feature encoding units of all layers is then

$$F = \sum_{\kappa=0}^{k} f_\kappa = (k+1)n - \left(n - \frac{n}{m}\right)\frac{k(k+1)}{2k} = \frac{n}{2}\frac{m+1}{m}(k+1).$$

Adding the links emanating upwards from all but the top-most layer, we get the total number of units as

$$\begin{aligned} N(n,k) &= F + \frac{1}{\alpha}\left(F - \frac{n}{m}\right)\ l \\ &= \frac{n_\alpha}{2}\frac{m+1}{m}\left[\alpha(k+1) + \left(k+1-\frac{2}{m+1}\right)\left(\frac{n_\alpha}{m}\right)^{\frac{1}{k}}\right]. \end{aligned} \tag{3.2}$$

Setting the derivate with respect to k to zero leads to

$$-\alpha\left(\frac{n_\alpha}{m}\right)^{-\frac{1}{k}} = 1 + \frac{\ln\frac{n_\alpha}{m}}{k^2}\left(\frac{2}{m+1} - k - 1\right).$$

With the ansatz

$$k_{opt} = \tilde{c}\ln\frac{n_\alpha}{m} \tag{3.3}$$

this turns into

$$-\alpha e^{-\frac{1}{\tilde{c}}} = 1 + \frac{1}{\tilde{c}^2\ln\frac{n_\alpha}{m}}\left(\frac{2}{m+1} - 1\right) - \frac{1}{\tilde{c}}.$$

For large input/output ratio m, the term $\frac{2}{m+1}$ becomes negligible, so that $\tilde{c}$ depends only on the number of independently routed output nodes $\frac{n_\alpha}{m}$ and not on m itself:

$$-\alpha e^{-\frac{1}{\tilde{c}}} \approx 1 - \frac{1}{\tilde{c}^2\ln\frac{n_\alpha}{m}} - \frac{1}{\tilde{c}}. \tag{3.4}$$

Numerical analysis of (3.4) shows, however, that $\tilde{c}$ changes by less than 10% when $\frac{n_\alpha}{m}$ is varied over 3 orders of magnitude. So we can say that, like c in Sect. 2, $\tilde{c}$ only depends on the parameter α. Fig. 3 shows that $\tilde{c}$ takes on similar but slightly higher values than c.

Calculating the size of the derived routing circuit by plugging (3.3) into (3.2) yields

$$N_{opt} = \left(\alpha + e^{\frac{1}{\tilde{c}}}\right)\frac{n_\alpha}{2}\left(\tilde{c}\ln\frac{n_\alpha}{m} + 1\right) \tag{3.5}$$

for large m. Although the relation is a bit different from the one derived for equal input and output domains (2.8), the scaling with $\tilde{n_\alpha}\ln\tilde{n_\alpha}$ (here $\tilde{n_\alpha} = \frac{n}{\alpha m}$) remains the same.

4 Physiological interpretation

The main goal of this paper is to raise the question of optimal information routing in terms of required neural resources. In Sects. 2 and 3 we found that the optimal number of link layers in a routing circuit is given by

$$k_{opt} = c \ \ln n_\alpha$$

and

$$k_{opt} = \tilde{c} \ln \frac{n_\alpha}{m}$$

for routing to an output stage of identical size and of much smaller size, respectively. So in both cases, k_{opt} is proportional to the natural logarithm of the number of independently routed *output* nodes. The well defined prefactors c and $\tilde{c}$ are very similar (cf. Fig. 3), depend only on α, and do not vary too much over large ranges of α.

How do those results match the facts in the human brain? A good starting point is the optic nerve, which is known to contain $\sim 10^6$ fibers for humans and other primates (Potts, Hodges, Shelman, Fritz, Levy and Mangnall, 1972). Since the optic nerve is the bandwidth bottleneck of the visual system, it is safe to assume that it contains no redundant information. The number of neurons in V1, however, is by far higher than the number of optic nerve fibers, mainly for two reasons. First, the cortex employs a population coding strategy in order to reduce noise and increase transmission speed. This means that several neurons together (perhaps the ≈ 100 of a cortical minicolumn) represent one of our abstract feature units. Second, visual information is represented in an overcomplete code in V1 (Olshausen and Field, 1997), increasing the number of feature units over the number of optic nerve fibers. Nevertheless, the information represented in V1 cannot be higher than that transported by the optic nerve, so that overcomplete groups of units can be routed collectively. We will therefore assume that the number of feature encoding units is of the same order as the number of fibers in the optic nerve, keeping in mind that an overcomplete basis in V1 may be accounted for by a correspondingly higher value of α.

From the primary visual area V1, visual information is routed retinotopically along the ventral pathway to a target region in inferotemporal cortex (IT). Psychophysical evidence (van Essen, Olshausen, Anderson and Gallant, 1991) suggests that about 1000 feature nodes are sufficient to represent the contents of the two dimensional "window of attention", and therefore

the size of this target region, at any given time. One may assume that there exist multiple such target regions in parallel in IT, which are used for different object recognition tasks. But that question is outside the scope of this paper.

How would our routing architecture look for these numbers? For this, we still miss an estimate of the parameter α. Research in our lab has shown that representing an image with 40 Gabor wavelets in each image point preserves all necessary image information of gray scale images (Wundrich, von der Malsburg and Würtz, 2004) and allows good object identification (Lades, Vorbrüggen, Buhmann, Lange, von der Malsburg, Würtz and Konen, 1993). To additionally include color and temporal information (direction of motion), this number would have to be roughly twice as high. This is in line with findings concerning the number of orientation pinwheels in the primate brain. (Obermayer and Blasdel, 1997) report around 10^4 pinwheels for V1 of the Macaque. Assuming a similar number for the human brain, we face the situation of an input region of the ventral stream containing 10^6 feature units clustered in some 10^4 pinwheels. If we assume that every pinwheel—as a first order approximation of the functional "hypercolumn"—contains the full set of visual features for a certain input location on one retina, it follows that the number of these distinct features is of the order 100. Inputs from the two eyes are treated independently here, so that successful stereoscopic fusion can be achieved for arbitrary depths by activating the right routing links.

While we have two agreeing estimates of the number of feature units per resolution point, coming from computer vision and physiology, the number α of features that can be routed together is difficult to estimate. It depends on the kinds of invariance operations that are actually realized in the routing circuit, as discussed in Sect. 2.1. We assume α to lie in the approximate range of 2 to 5. For these values, the optimal number of layers for routing $(k+1)$ from a 10^6 node input to a 1000 node output ranges from 4.3 to 5.8. Fig. 5 shows these values, as well as the number of units required for the full circuit when using the optimal number of layers.

The ventral pathway comprises the areas V1, V2, V4, and IT. IT in turn consists of posterior, central, and anterior parts. In our setting it may make sense to take into account this additional subdivision, since the receptive field sizes of these three parts are very different ((Tanaka, Fujita, Kobatake, Cheng and Ito, 1993), see also Fig. 4 in (Oram and Perret, 1994)), suggesting that they form different stages of the routing hierarchy. Visual information

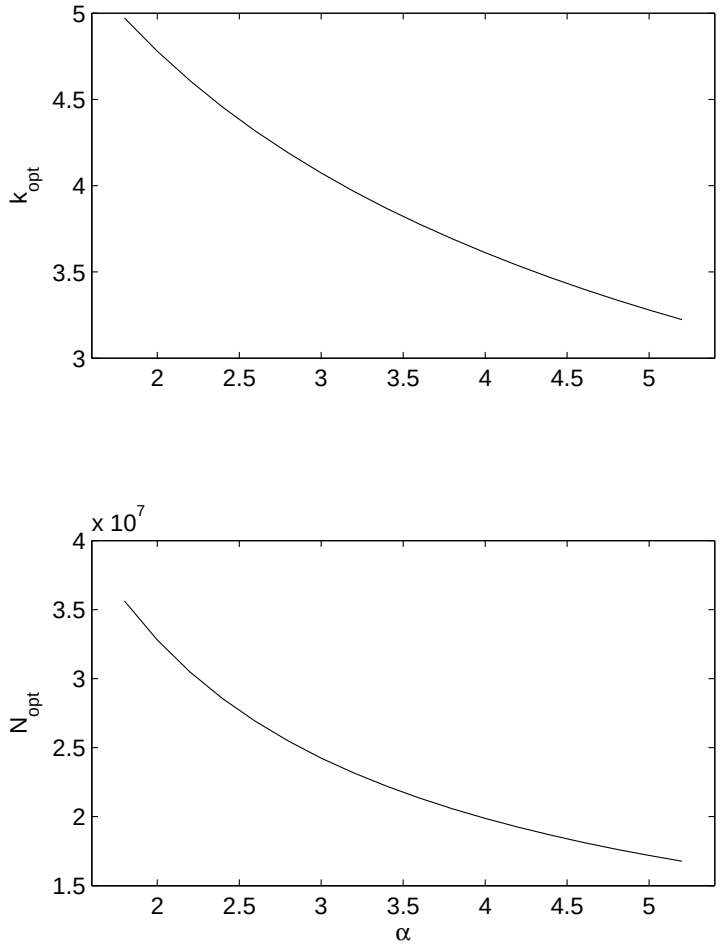


Figure 5: Routing from an input stage of 10^6 to an output stage of 1000 nodes. The upper part displays the optimal number of link layers as a function of α. Below we see the total number of required units using the optimal number of layers from above.

is relayed from the lateral geniculate nucleus in a rather clear sequential order V1→V2→V4→PIT→CIT→AIT, finally being combined with other signal streams in the superior temporal polysensory area (STP). Note that there exist at least equally strong feed*back* connections between the layers, indicating the importance of recurrent processes in vision. The number of 4–6 distinct cortical stages (depending on whether we regard IT as one or three stages) lies clearly in the range derived for our optimal circuit above. It is therefore possible that the ventral pathway indeed performs computationally optimal information routing. At this point, however, this is only a hypothesis, due to the great uncertainties in the available data. More explicit interpretations would be possible if α could be narrowed down further (also by quantifying the "overcompleteness" of V1) and if the stages involved in the routing were known for certain. Also, more psychophysical work on the information content in the window of attention would be desirable.

5 Discussion

We have seen in Sects. 2 and 3 that under some very general assumptions there exists a clear optimality condition on the number of layers required to build a routing architecture with minimal neural resources. This number depends on the size of the target region as well as the number of independently routed feature types. Within the given uncertainties, the derived numbers agree well with physiological data.

Constraining the design of a routing architecture by an optimality condition, as we did, has the advantage of imposing an additional requirement to an otherwise underconstrained problem. While the Shifter Circuit of (Olshausen et al., 1993) addresses several anatomical and physiological facts, there is no experimental or theoretical justification for some of the parameter values chosen, among them the exact doubling of link spacing from layer to layer. In the absence of experimental results dictating these values, we think it best to follow some global optimality condition like proposed above.

We are well aware that our very general assumptions can be refined in several ways, possibly changing the derived quantitative results:

- We avoided on purpose a detailed discussion of the kind of feature-to-feature connectivity that may be in place to achieve scale and orientation invariance. This will be addressed in future work and will help to narrow down the parameter α.

- Routing architectures with many numbers of layers are a disadvantage to the organism in terms of longer reaction times and more complicated routing dynamics. This additional cost has not been considered here, its influence would bias the biological routing architecture in favor of fewer stages than derived here.

Although the main contribution of this communication may lie in introducing optimality criteria to the design of routing circuits, it also leads to experimental predictions. One such prediction arises from the fact that the notion of a static receptive field becomes meaningless if one embraces an active routing process. During attention focusing and recognition, this process would choose a certain routing path and deactivate all alternative pathways. For a unit at the output stage in the hierarchy (IT), this would change the functional receptive field from a very broad region to a narrow and specific location. A unit at a medium stage of the hierarchy might even be bypassed by the currently established routing pathway. There is ample evidence for the behavioral plasticity of receptive fields (Moran and Desimone, 1985; Connor, Gallant, Preddie and van Essen, 1993), and recent findings (Murray, Boyaci and Kersten, 2006) show that even the size of representation in V1 can change with an object's perceived size (suggesting a scale invariant routing process that already starts in the mapping from LGN to V1). However, these findings are often interpreted as the result of a diffuse "attention modulation" mechanism, without taking the possibility of an explicit routing process seriously. In the light of the rather specific geometric changes of receptive fields implied by the presence of such a process, it should be possible to design attention experiments that can clearly prove or refute the routing hypothesis.

While the above predictions are general implications of an active routing process and have been discussed similarly before (Olshausen et al., 1993), the quantitative results obtained here make some more specific predictions. An interesting feature of the minimal architecture, already mentioned in Sect. 2 is that the number of links emanating from one node (Eq. (2.1) or (3.1)) is independent of network size:

$$l_{\mathrm{opt}} = n_\alpha^{\frac{1}{k_{\mathrm{opt}}}} = \exp\left(\frac{\ln n_\alpha}{\tilde{c}\ln n_\alpha}\right) = \exp\left(\frac{1}{\tilde{c}}\right). \tag{5.1}$$

l_{opt} is surprisingly low (between 3 and 9 for the range of α shown in Fig. 3). This number should not be confused with the full number of connections that a cortical neuron makes, which is known to be several thousand. First,

here we only count the connections necessary for information routing, not those involved in other kinds of processing or communication. Second, as mentioned above, the functional units discussed here are abstract "image points", which in the cortex are probably made up of ≈ 100 spiking neurons (like a cortical minicolumn). Single neurons in such a group would have to devote the majority of their connections to homeostatic within-group connections (Lücke and von der Malsburg, 2004), which do not appear on our level of abstraction. Nevertheless, the small fanout necessary for optimal routing is an interesting feature and shows that by including the number of control units into our optimization we have implicitly also minimized the required connectivity of the routing architecture.

The optimal number of layers (Eq. (2.4') or (3.3)), on the other hand, scales logarithmically with network size:

$$k_{opt} = \tilde{c} \ln n_\alpha .$$

This means that if more visual information has to be routed, the number of routing stages increases, while the local properties (number of connections that each node has to make) remain the same. Consequently, for species processing different amounts of visual information, the ventral streams should contain different numbers of routing stages. While the optic nerve of primates contains on the order of 10^6 fibers (Potts et al., 1972), the number is 10^5 for the rat (Fukuda, Sugimoto and Shirokawa, 1982), $2 \cdot 10^5$ for the cat (Hughes and Wässle, 1976), and $2.4 \cdot 10^6$ for the adult chicken (Rager and Rager, 1978). If we assume, as we did before, that the number of optic nerve fibers is a measure for the number of input units of the ventral stream, *and* if the number of output (IT) units changes by the same factor, then a rat would optimally have 2.3 layers less, a cat 1.6 layers less, and a chicken 0.9 routing layers more than a primate. The differences might be smaller, however, if the size of the output stage does not change as strongly as the number of optic nerve fibers, since k_{opt} depends on the number of output units. Although anatomical comparisons across species will be difficult, it may be interesting to investigate different brains with regard to this question.

6 Acknowledgments

We thank Charles Anderson for inspiring discussions, Cornelius Weber for useful comments on the manuscript, and two anonymous reviewers whose

criticism has helped us to clarify a number of points. This work was supported by the European Union through project FP6-2005-015803 ("Daisy") and by the Hertie Foundation.

References

Connor, C. E., Gallant, J. L., Preddie, D. C. and van Essen, D. C.: 1993, Responses in area v4 depend on the spatial relationship between stimulus and attention, *Journal of Neurophysiology* **75**, 1306–1308.

Cooley, J. W. and Connor, J. W. C. E.: 1965, An algorithm for the machine calculation of complex fourier series, *Mathematics of Computation* **19**, 297–301.

Cox, D., Meier, P., Oertelt, N. and DiCarlo, J. J.: 2005, 'breaking' position-invariant object recognition, *Nature Neuroscience* **8**(9), 1145–1147.

Dill, M. and Fahle, M.: 1998, Limited translation invariance of human visual pattern recognition., *Perception and Psychophysics* **60**(1), 65–81.

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J. and Wandell, B. A.: 2003, Visual field representations and locations of visual areas v1/2/3 in human visual cortex, *Journal of Vision* **3**, 586–598.

Fukuda, Y., Sugimoto, T. and Shirokawa, T.: 1982, Strain differences in quantitative analysis of rat optic nerve, *Experimental neurology* **75**, 525–532.

Hughes, A. and Wässle, H.: 1976, The cat optic nerve: Fibre total count and diameter spectrum, *The Journal of Comparative Neurology* **169**, 171–184.

Lades, M., Vorbrüggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. and Konen, W.: 1993, Distortion invariant object recognition in the dynamic link architecture, *IEEE Transactions on computers* **42**, 300–311.

Lücke, J.: 2005, Dynamics of cortical columns – sensitive decision making, *Proc. ICANN*, LNCS 3696, Springer, pp. 25 – 30.

Lücke, J. and von der Malsburg, C.: 2004, Rapid processing and unsupervised learning in a model of the cortical macrocolumn, *Neural Computation* **16**, 501 – 533.

Moran, J. and Desimone, R.: 1985, Selective attention gates visual processing in the extrastriate cortex, *Science* **229**, 782–784.

Murray, S. O., Boyaci, H. and Kersten, D.: 2006, The representation of perceived angular size in human primary visual cortex, *Nature Neuroscience* **9**, 429–434.

Obermayer, K. and Blasdel, G. G.: 1997, Singularities in primate orientation maps, *Neural Computation* **9**, 555–575.

Olshausen, B. A. and Field, D. J.: 1997, Sparse coding with an overcomplete basis set: a strategy employed by v1?, *Vision Research* **37**, 3311–3325.

Olshausen, B. A., Anderson, C. H. and van Essen, D. C.: 1993, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *Journal of Neuroscience* **13**(11), 4700–4719.

Oram, M. W. and Perret, D. I.: 1994, Modeling visual recognition from neurobiological constraints, *Neural Networks* **7**, 945–972.

Perkel, D. J., Bullier, J. and Kennedy, H.: 1986, Topography of the afferent connectivity of area 17 in the macaque monkey: A double-labelling study, *The Journal of Comparative Neurology* **253**, 374–402.

Pitts, W. and McCulloch, W. S.: 1947, How we know universals: the perception of auditory and visual forms, *Bulletin of Mathematical Biophysics* **9**, 127–147.

Postma, E. O., van den Herik, H. J. and Hudson, P. T. W.: 1997, Scan: A scalable neural model of covert attention, *Neural Networks* **10**(6), 993–1015.

Potts, A. M., Hodges, D., Shelman, C. B., Fritz, K. J., Levy, N. S. and Mangnall, Y.: 1972, Morphology of the primate optic nerve. i. method and total fiber count, *Investigative Ophthalmology & Visual Science* **11**, 980–988.

Rager, G. and Rager, U.: 1978, Systems-matching by degeneration i. a quantitative electron microscopic study of the generation and degeneration of retinal ganglion cells in the chicken, *Experimental Brain Research* **33**, 65–78.

Schwartz, E. L.: 1977, Spatial mapping in primate sensory projection: analytic structure and relevance to perception., *Biological Cybernetics* **25**, 181–194.

Tanaka, K., Fujita, I., Kobatake, E., Cheng, K. and Ito, M.: 1993, Serial processing of visual object-features in the posterior and anterior parts of the inferotemporal cortex, *in* T. Ono, L. R. Squire, M. E. Raichle, D. Perrett and M. Fukuda (eds), *Brain mechanisms of perception and memory: From neuron to behaviour*, Oxford University Press, pp. 34–46.

Tanigawa, H., Wang, Q. and Fujita, I.: 2005, Organization of horizontal axons in the inferior temporal cortex and primary visual cortex of the macaque monkey, *Cerebral Cortex* **15**, 1887–1899.

van Essen, D. C., Olshausen, B., Anderson, C. H. and Gallant, J.: 1991, Pattern recognition, attention, and information bottlenecks in the primate visual system, *in* B. Mathur and C. Koch (eds), *Proceedings of the SPIE Conference on Visual Information Processing: from neurons to chips*, Vol. 1473, pp. 17–28.

Wundrich, I. J., von der Malsburg, C. and Würtz, R. P.: 2004, Image representation by complex cell responses., *Neural Computation* **16**(12), 2563–2575.

Zhu, J. and von der Malsburg, C.: 2004, Maplets for correspondence-based object recognition., *Neural Networks* **17**(8-9), 1311–1326.