Visual information processing

How Does the Brain Solve Visual Object Recognition? [DiCarlo et al., 2012]
Integrated model of visual processing [Bullier, 2001]
Vision as a feed-forward problem.

Vision operates on different time scales and facilitates many functions:
- object recognition, object tracking, segmentation, obstacle avoidance, object grasping

Core Object Recognition:
- fast first recognition phase. \( \approx 250 \text{ms} \) in monkey, \( \approx 350 \text{ms} \) in human and can be presented at 100 ms rate.
- in agreement with rapid eye movement during exploration where the eye is fixated 200-500 ms.

Humans and monkeys perform core recognition fast, despite invariance problem.
Vision as a feed-forward problem.

Early visual areas (retina, LGN, V1) have tangled representation of objects.

Higher areas (IT) untangle into ’linear separable’ representation.

Vast majority of images are ’unlabelled’ but temporally contiguous, which may facilitate unsupervised learning.

’Shape’ results as only feature that is not temporally contiguous
Vision as a feed-forward problem.

Ventral stream houses key circuits that underlie object recognition:
- lesions in the posterior ventral stream produce complete blindness in part of the visual field
- lesions or inactivation of anterior regions, (IT), can produce selective deficits in the ability to distinguish among complex objects
- retinotopic mappings for V1, V2, V4. less clear in IT.
Feed-forward model

Dominant model of visual processing is feed-forward [Marr, 1982].
- first local information processing in V1, V2
- later global information processing in higher layers (V4, IT)
- for instance orientation selectivity in V1 is computed from LGN center-surround input (Hubel Wiesel)

Problem:
- feed-forward model fails when input is noisy, while humans succeed (Mumford 1993)
- reason is recognition of local features fails and cannot be fixed by global integration
- weakness is that local ambiguity cannot be resolved from global feed-back.
This paper argues that:

Cortical processing requires that information be exchanged between neurons coding for distant regions in the visual field.

It is argued that feedback rather than lateral connections are the best candidates for such rapid long-distance interconnections.
Lateral connectivity

Early areas such as V1, V2 have small receptive fields and high magnification. Therefore horizontal connections only reach short distance:
- the axon of a V1 or V2 neuron with a foveal receptive field cannot reach beyond $0.6^\circ$ (assuming a maximal axon length of 6 mm).

Higher areas have larger receptive fields and smaller magnification. These areas encode specialised features (movement direction, depth, color, shape).
- they can spatially integrate their particular feature, integration across features much more difficult

Feedback:
- feedback connections have very large convergence regions. can carry information from long distances in the visual field (Salin 1993)

Large scale spatial integration by
- lateral connectivity in higher layers. Requires (fast) computation combining different features.
- feed-back from higher layers to lower layers.
Feed-back connectivity

90% of the information transferred by a V1, V2 neuron is transmitted during the first 100 ms of its response to a visual stimulus. Thus feed-back requires:

- rapid activation of higher order neurons
- fast feed-back connections

Areas MT, MST, FEF (frontal eye field) and 7a that contain neurons that are activated sufficiently early to influence neurons in areas V1 and V2.

Speed of feed-back is in order of 1-2 ms because axons between areas are myelinated (for comparison, speed of lateral connections is \( \approx 0.1 \text{m/s} \), requiring about 100 ms to reach neurons 1° apart in V1).
Magnocellular neurons of the LGN reach area V1 and beyond some 20 ms earlier than activity transferred from the parvocellular neurons of the LGN [34]. The early activation of the dorsal stream after visual stimulation (Fig. 1B) most likely results from its almost exclusive drive by the M channel [26] and the high degree of myelination in many areas of the ‘fast brain’.
Empirical evidence: human fMRI

Moving Kanisza rectangles create strong motion percept (left) absent in the control stimulus (right).

As expected, stimulus produces strong signal in V5/MT, and LOS/KO responsible for object motion. Surprisingly, also strong activity in V1, whose neurones have small receptive fields and are not activated by the stimulus. Probable interpretation is feedback from V5/MT to V1, V2 neurons.
Empirical evidence: neural activity in V3

Moving low contrast bar over background of white and gray rectangles. Conditions:
- moving bar on static background (centre)
- moving bar on co-moving background (centre + background)
- moving background (background)

Area MT is known to be mainly involved in low contrast:
- inactivation of MT removes feed-back in centre condition (A left)
- active and inactive MT yields suppression in centre + background condition (A centre)
Empirical evidence: neural activity in V3

Area MT is known to be mainly involved in low contrast:
- decrease of centre response due to MT cooling strongest at low saliency (B)
- background induced suppression (centre minus centre+background) almost eliminated at low saliency due to MT cooling (C)
Empirical evidence: stimulus flash

Effect of MT cooling on response of 51 V1,V2,V3 neurons shows that feedback is as fast as 10 ms.
References


