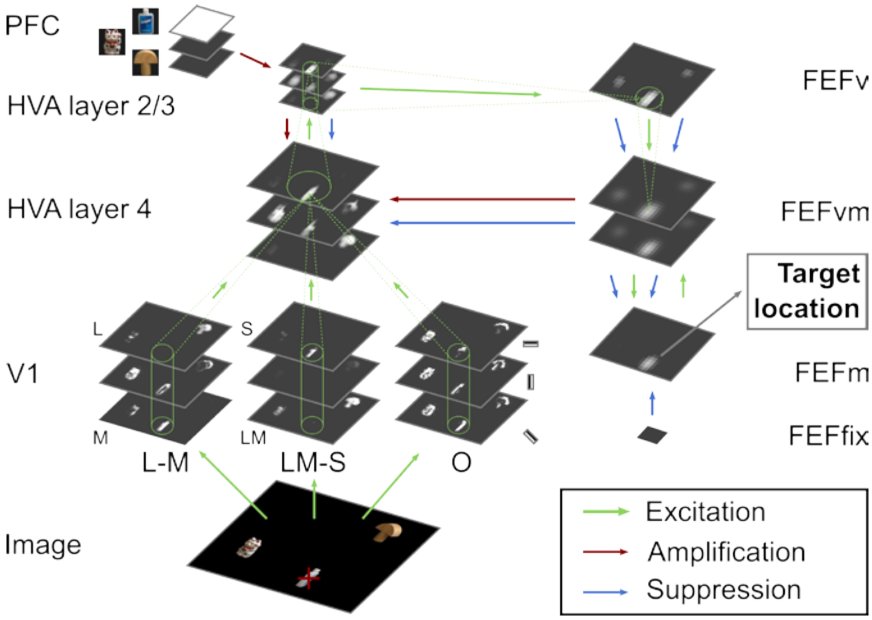

Figure 1 The novel model of visual attention. The processing is illustrated in a guided visual search task to localize and recognize the target 'bottle', indicated by the red cross. Firstly, the image is processed by a primary visual cortex model (V1), encoding oriented edges (*O*), red-green (*L - M*), and blue-yellow color contrasts (*S - LM*). From this, a higher visual area (HVA), comparable to V4 or IT, recognizes object views. The task implies feature-based attention towards the target object, simulated by activating a target preferring neuron in the prefrontal cortex (PFC). This results in a feature-based amplification signal to HVA, increasing the responses of target view neurons in both HVA layers, and so selects the target. In parallel, feature-based suppression from layer 2/3 to layer 4 decreases neuronal responses towards distractors. The so modulated HVA responses provide an target specific bias to the frontal eye field (FEF) and so guide eye movements. The FEF forms together with HVA a recurrent loop, which focuses activity over time on a single target location, leading to the emergence of spatial attention. Finally, an eye movement is planned towards this location, indicating the outcome of the recognition process.

Extended abstract: Visual attention modulates neuronal activity in a rich set of paradigm (Reynolds & Heeger, 2009, Neuron). However these paradigms use simple stimuli and setups, making it hard to link their physiological findings to complex real-world tasks containing whole objects and background clutter. Here we want to bridge this gap in a real-world, guided visual search task, in which an object of interest has to be searched and recognized in a scene. Computer vision systems already use visual attention for such object recognition (Borji & Itti, 2013, IEEE TPAMI), but only as a pre-selection mechanism for a subsequent, sophisticated recognition stage (saliency map approach). Contrary, top-down attention in the primate brain seems to be a control network modulating neuronal activity for the current task (Miller & Buschman, 2013, Curr Opin Neurobiol). Thus attention does not serve as a separate pre-selection stage, instead it controls both selection and recognition processes. This approach is advantageous compared to computer vision systems as it solves the chicken-egg problem of segmentation and recognition by executing both processes in parallel (Antonelli et al., 2014, IEEE TAMD).

Here, we present a novel model of attention (Fig. 1) which uses the cortical processing of attention as a core principle, and underpins it with a large set of findings from physiology and neuroanatomy. A higher visual area (HVA) is simulated by a scaled version of a recently developed microcircuit model of visual attention (Beuth & Hamker, in revision, Vision Res), which explains the physiologically data of twelve different attention experiments, e.g. biased competition, modulation of contrast response functions, tuning curves, and surround suppression. It primarily relies on neuronal implementations of a few attention mechanisms: amplification, divisive normalization, spatial pooling, and suppression. The area contains view-tuned neurons to represent objects, learned by a trace-learning rule (Antonelli et al., 2014). The HVA is part of a larger visual cortex model which possesses neuroanatomical properties like top-down attentional processing (Miller & Buschman, 2013), hierarchical receptive field sizes (Smith et al., 2001, Cerebral Cortex), and synaptic transmission delays (Schmolensky et al., 2000, J Neurophysiol). We also included a model of the frontal eye field (Zirnsak et al., 2011, Eur J Neurosci) with its physiological cell types (Schall, 1991, J Neurophysiol): visual (FEFv), movement (FEFm) and visuomovement cells (FEFvm). To sum up, we developed a very general model of visual attention embedded in the visual and prefrontal cortices.

We evaluated the model on two large and realistic object recognition test sets, consisting of 1000 different scenes with either a) black, or b) white-noise backgrounds. Each scene contains five different objects from a set of 100 objects under 72 different rotations (COIL-100 data set). The model´s task was to search for one of these five objects and to report its location. The model achieves 92% accuracy on black backgrounds and 71% on white-noise backgrounds. In the black background set, misrecognitions occur mainly when a distractor is similar to the target and also more salient as the target. These cases occur more often in the white-noise set as the noise reduces the neuronal representation of targets. Additionally, misrecognitions occur in this set if an object is similar to the background noise.

We furthermore investigated the neuronal mechanisms of attention in the guided visual search task. We observe that feature-based amplification, explaining the multiplicative effect of feature-based attention at the neuronal activity, selects the target by amplifying target specific neurons. Deactivating feature-based amplification leads to a selection of the most salient target. Interestingly, the white-noise backgrounds require a stronger amplification signal as the black backgrounds because amplification has to compensate for the reduced target activity. We also investigated feature-based suppression which accounts for decreased neuronal activity like in biased competition. In this task, it suppresses the neuronal activity resulting from distractors and from the background. Deactivating this mechanism increases such activity by a factor of 2.7 at the black background set, and by a factor of 2.0 at the white-noise set.