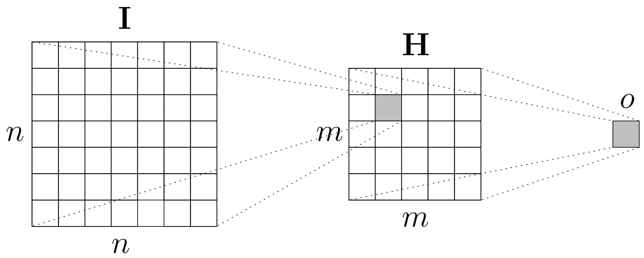
2 Artificial Intelligence (SBH)

Evil Robot is updating his visual system. He has a single camera that produces an $n \times n$ matrix $\mathbf{I}$ of pixel values. His visual system is arranged as follows:



The input $\mathbf{I}$ is reduced to an $m \times m$ matrix $\mathbf{H}(\mathbf{I})$. The elements $H_{i,j}$ are

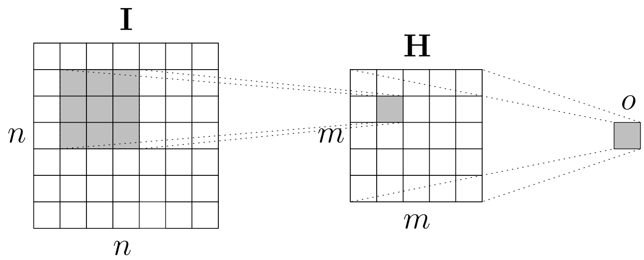
$$H_{i,j}(\mathbf{I}) = \sigma\left(\sum_{k=1}^{n}\sum_{l=1}^{n} w_{k,l}^{(i,j)} I_{k,l} + b^{(i,j)}\right)$$

where σ is an appropriate function, and $w_{k,l}^{(i,j)}$ and $b^{(i,j)}$ are the weights and bias for element (i,j). A single output $o(\mathbf{H})$ is computed as

$$o(\mathbf{H}) = \sigma\left(\sum_{k=1}^{m}\sum_{l=1}^{m} w_{k,l} H_{k,l} + b\right).$$

(*a*) If Evil Robot has a training example $(\mathbf{I}', y')$ and is using an error $E(\mathbf{w})$ where $\mathbf{w}$ is a vector of all weights and biases available, derive an algorithm for computing $\frac{\partial E}{\partial \mathbf{w}}$ for the example. [12 marks]

(*b*) A modification to the system works as follows:



The mapping from $\mathbf{I}$ to $\mathbf{H}$ is replaced by an $n' \times n'$ *convolution kernel*. This has a single set of parameters $v_{k,l}$ and c used to compute every element of $\mathbf{H}$ as the weighted sum of a patch of elements in $\mathbf{I}$

$$H_{i,j}(\mathbf{I}) = \sigma\left(\sum_{k=1}^{n'}\sum_{l=1}^{n'} v_{k,l} I_{i+k-1,j+l-1} + c\right).$$

Provide a detailed description of how the algorithm derived in Part (*a*) must be updated to take account of this modification. [8 marks]