

1 Convolutional layer

We propose two designs to imitate the operations of a convolutional layer. The key difference is that kernels may need to be cycled through the layers.

- **Encoded blur:** An amplitude mask with the flipped convolution kernels is placed at the aperture plane. At some defocus distance Δz from the conjugate image plane, ignoring diffraction effects, the image will be approximately convolved with the shape of the aperture. Cons: light loss; pros: color coding.
- **Phase mask:** A phase mask designed with point spread function (PSF) corresponding to the flipped is placed at the aperture plane. The image will be convolved with the phase mask PSF at the image plane. Cons: don't have full phase and amplitude control; pros: no light loss.

In both cases, the PSF may need to be extended to account for non-circular convolution, and we only keep the valid part of the convolution. Suppose the input image plane, I_{in} is an $n \times m$ array of same size images ($\{I_{(1,1)}, \dots, I_{(n,m)}\}$), with centers shifted by δ . The PSF is composed of an $n' \times m'$ flipped convolution kernels, $\{w_{(1,1)}, \dots, w_{(n',m')}\}$. For convenience, let negative indices of k and l circle back around (cyclic...). The image formation can be written

$$I_{\text{out}}(i, j) = (I_{\text{in}} * PSF)(i, j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{\text{in}}(k, l) \cdot PSF(i - k, j - l) dl dk \quad (1)$$

If there is sufficient padding around the sub-images for non-overlapping convolutions,

$$I_{\text{out}}(i, j) = \sum_{\eta=1}^n \sum_{\mu=1}^m \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} I_{(\eta,\mu)}(k, l) \cdot w_{(\eta,\mu)}(i - k, j - l) dl dk \quad (2)$$

In other words, the output image, I_{out} , is an $n' \times m'$ array of images, where

$$I_{\text{out},(\eta',\mu')}(i, j) = \sum_{\eta=1}^n \sum_{\mu=1}^m \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} I_{(\eta,\mu)}(k, l) \cdot w_{(\eta'-\eta,\mu'-\mu)}(i - k, j - l) dl dk \quad (3)$$

For a single input image, i.e. for the first convolutional layer, $n = m = 1$, so this simply produces multiple copies of the input image convolved with different PSFs.

$$I_{\text{out},(\eta',\mu')}(i, j) = \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} I_{\text{in}}(k, l) \cdot w_{(\eta'-\eta,\mu'-\mu)}(i - k, j - l) dl dk \quad (4)$$

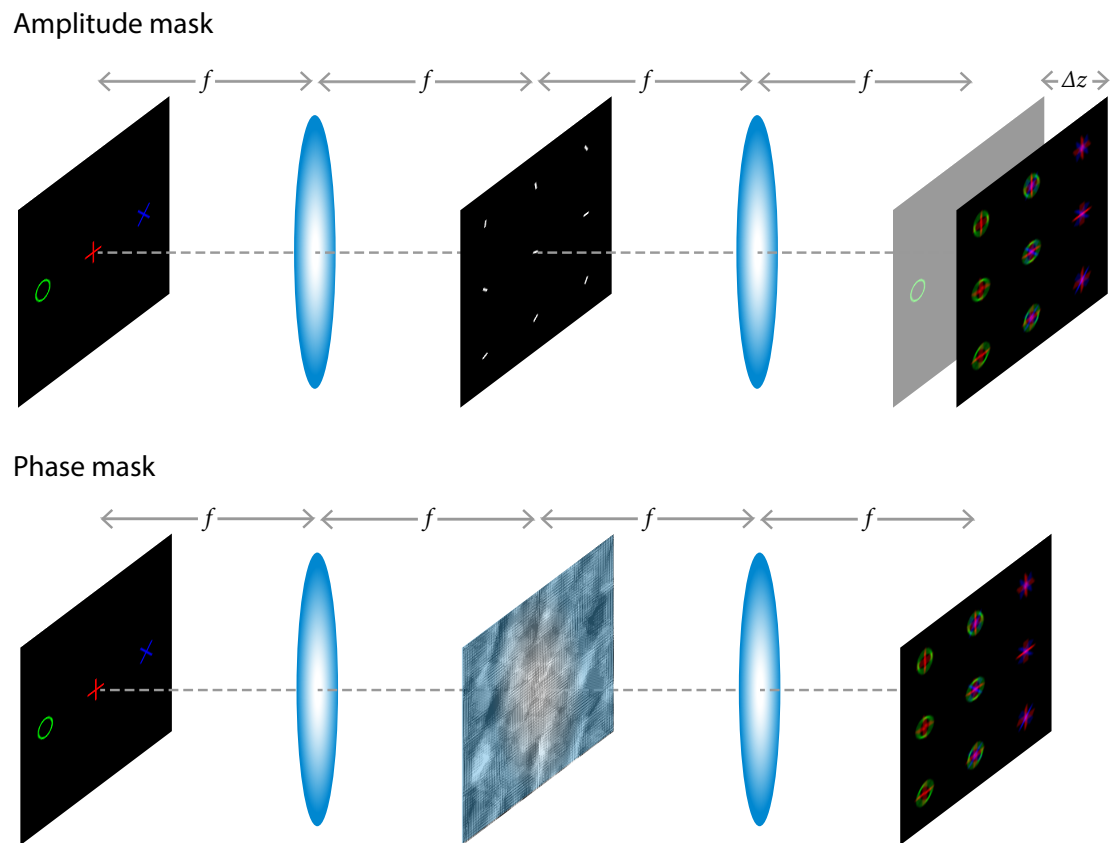


Figure 1: Optical designs for convolutional layer.