DEEP LEARNING

Practical Notes on Convolutional Neural Networks

Sigurd Spieckermann

May 7, 2015

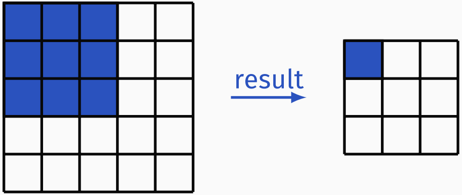
Research Scientist @ Siemens Corporate Technology

OVERVIEW

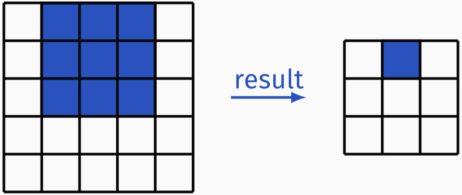
- Convolution Algorithms
- Reducing Overfitting
- Architecture Guidelines
- Kaggle: The Galaxy Zoo Challenge

CONVOLUTION ALGORITHMS

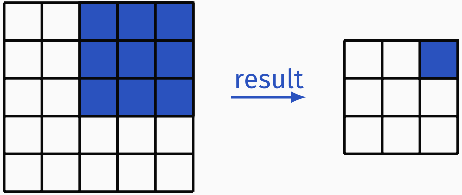
MODES: VALID CONVOLUTION



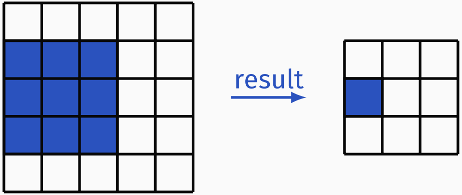
MODES: VALID CONVOLUTION



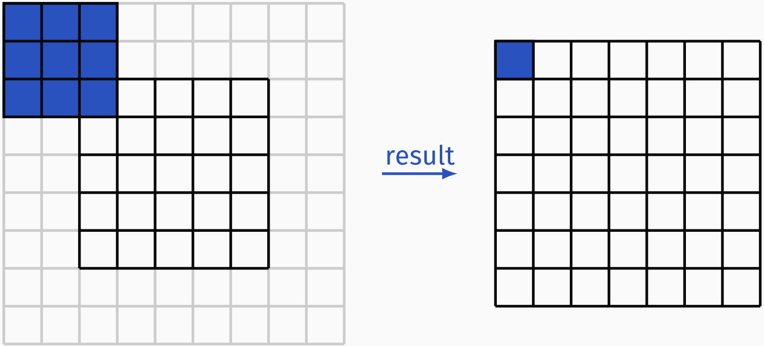
MODES: VALID CONVOLUTION



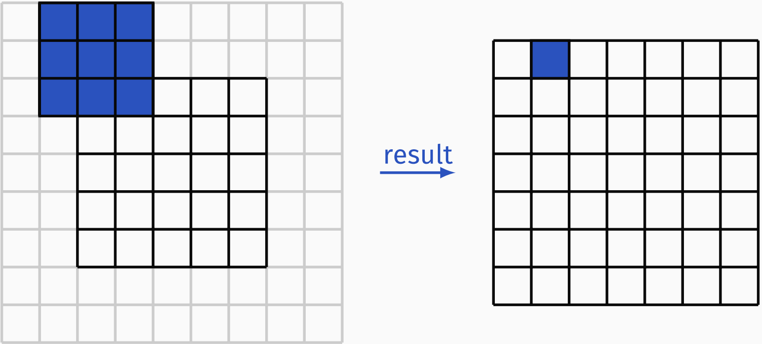
MODES: VALID CONVOLUTION



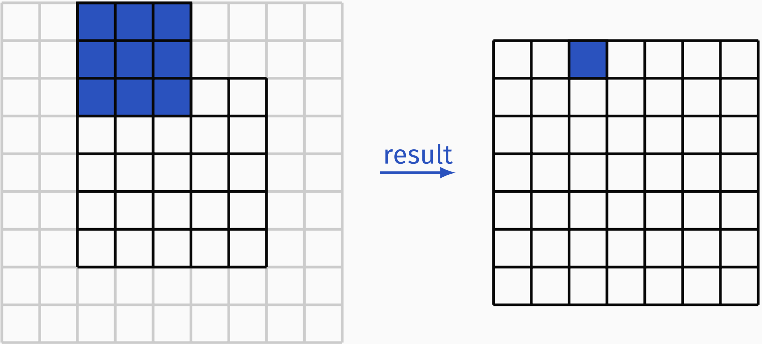
MODES: FULL CONVOLUTION



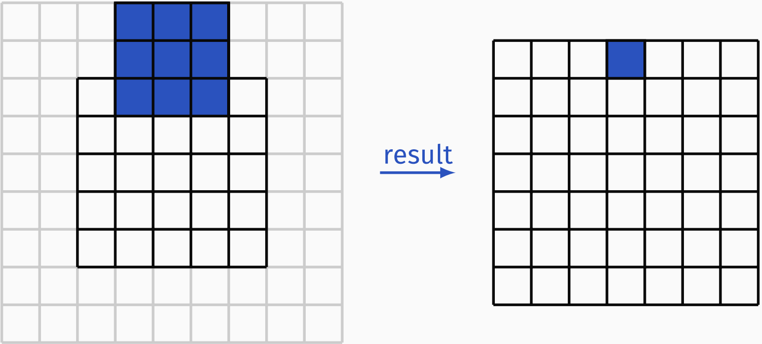
MODES: FULL CONVOLUTION



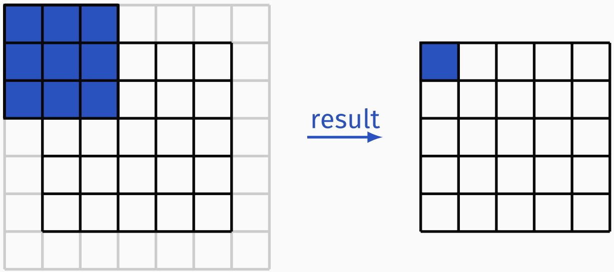
MODES: FULL CONVOLUTION



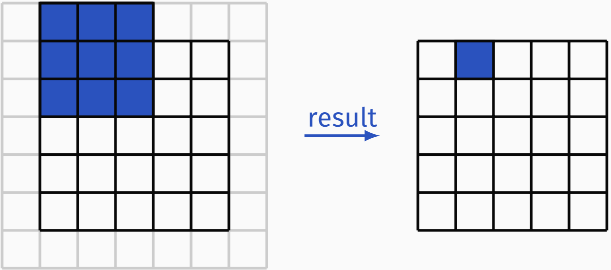
MODES: FULL CONVOLUTION



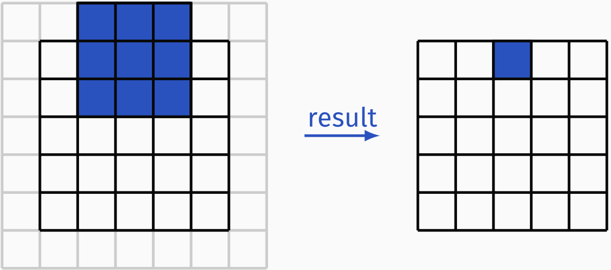
MODES: SAME CONVOLUTION



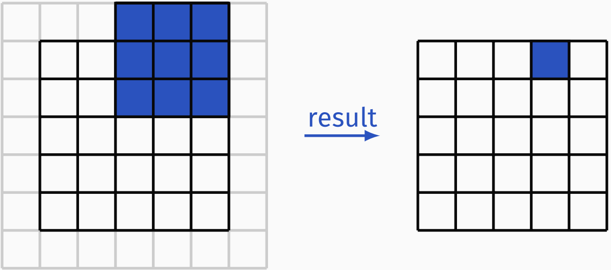
MODES: SAME CONVOLUTION



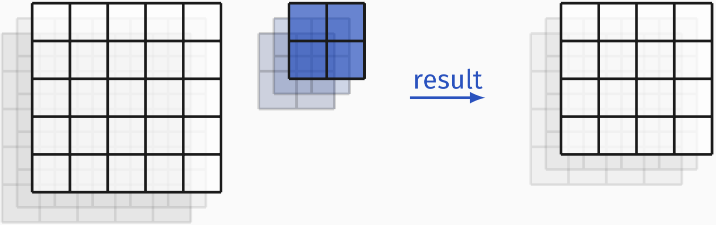
MODES: SAME CONVOLUTION



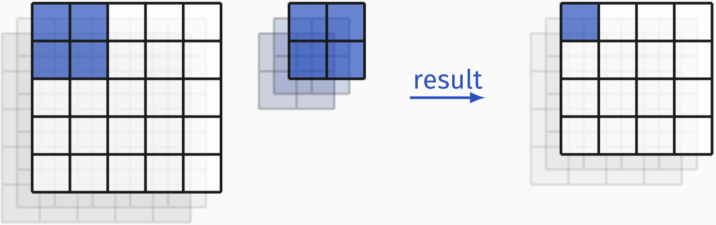
MODES: SAME CONVOLUTION



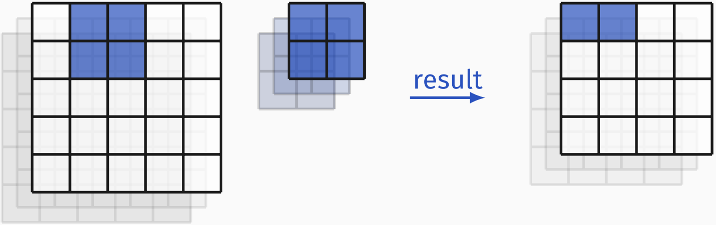
DIRECT CONVOLUTION



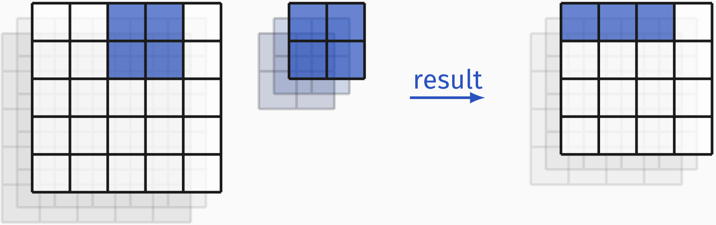
DIRECT CONVOLUTION



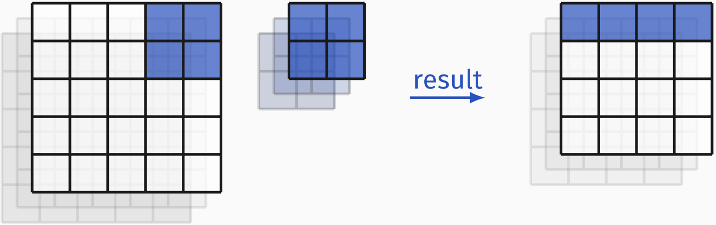
DIRECT CONVOLUTION



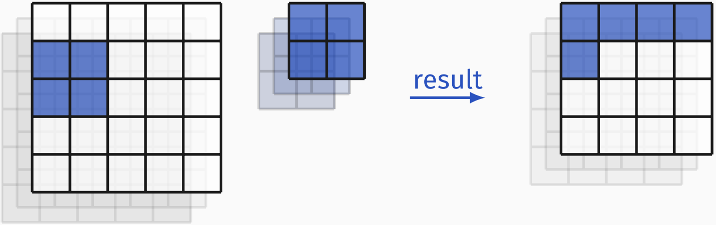
DIRECT CONVOLUTION



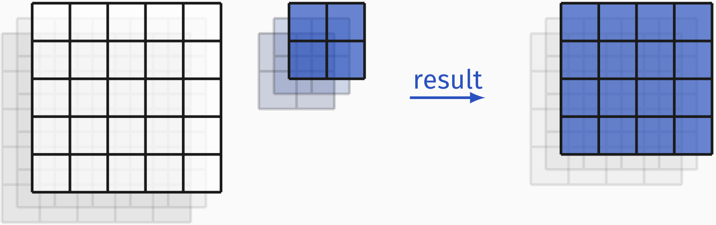
DIRECT CONVOLUTION



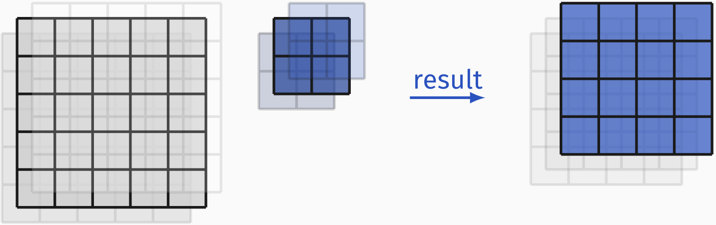
DIRECT CONVOLUTION



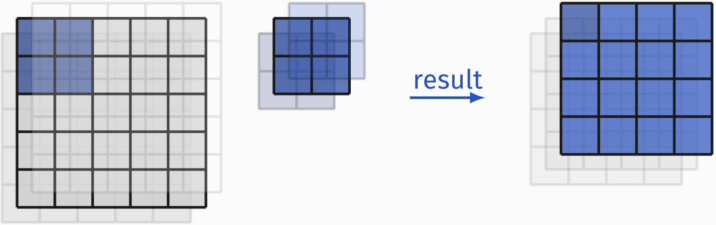
DIRECT CONVOLUTION



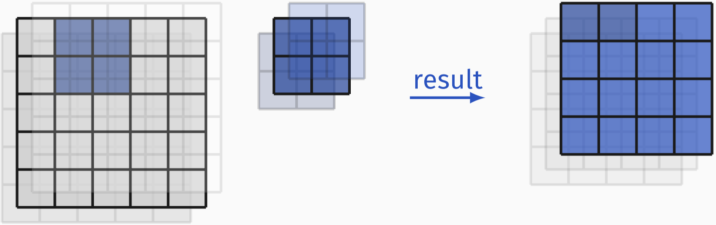
DIRECT CONVOLUTION



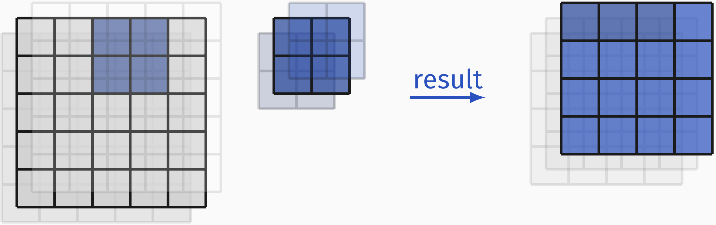
DIRECT CONVOLUTION



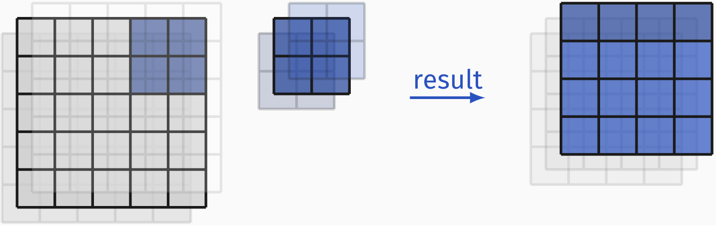
DIRECT CONVOLUTION



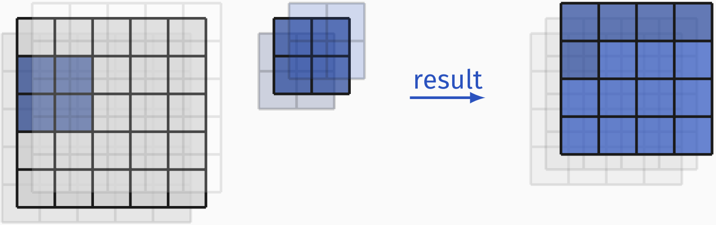
DIRECT CONVOLUTION



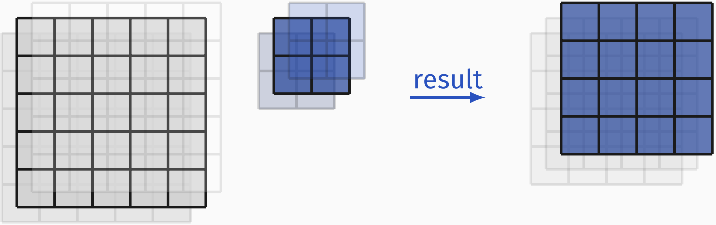
DIRECT CONVOLUTION



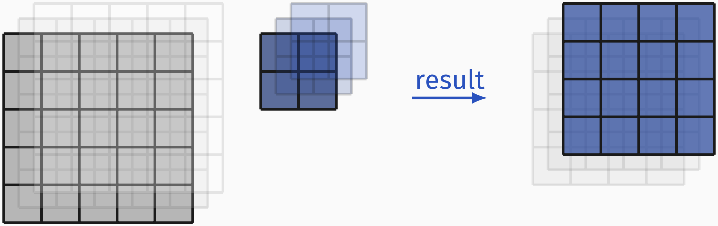
DIRECT CONVOLUTION



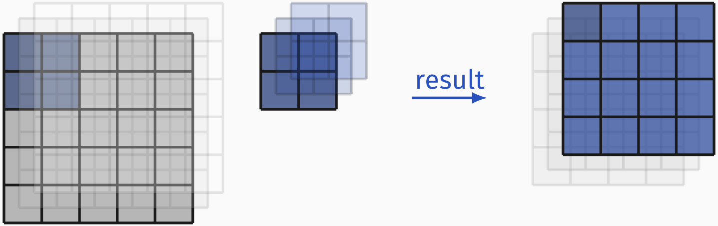
DIRECT CONVOLUTION



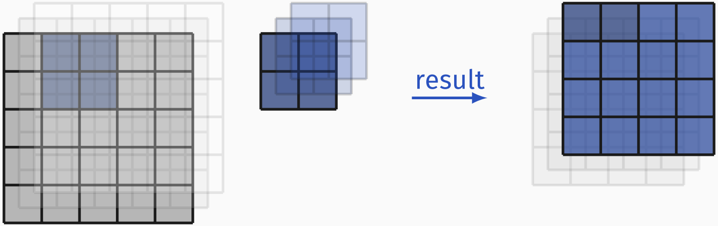
DIRECT CONVOLUTION



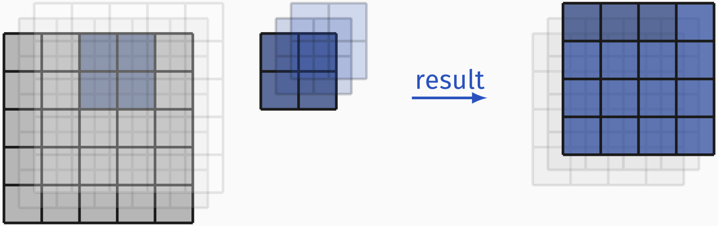
DIRECT CONVOLUTION



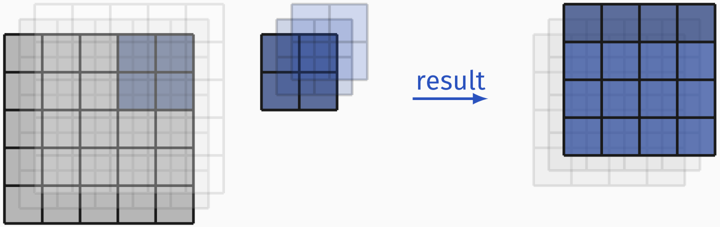
DIRECT CONVOLUTION



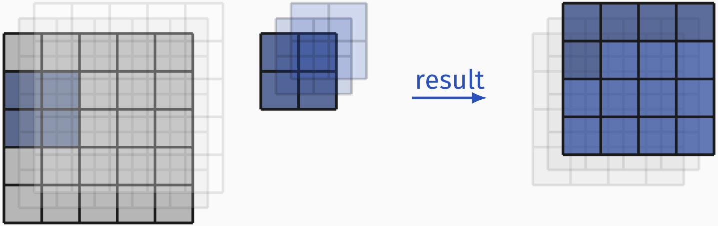
DIRECT CONVOLUTION



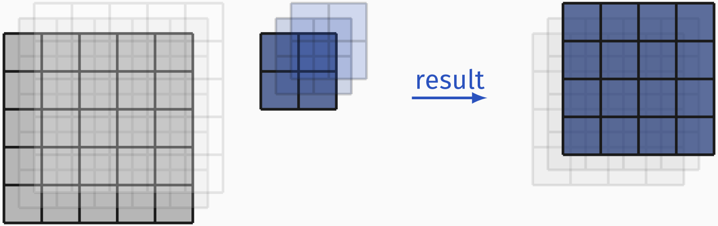
DIRECT CONVOLUTION



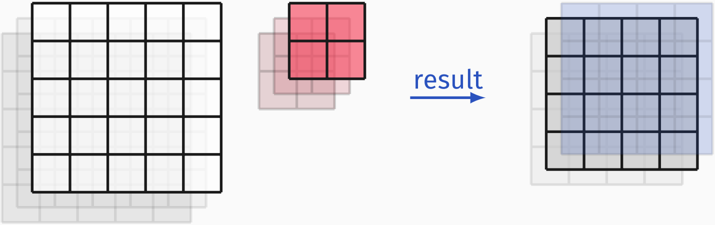
DIRECT CONVOLUTION



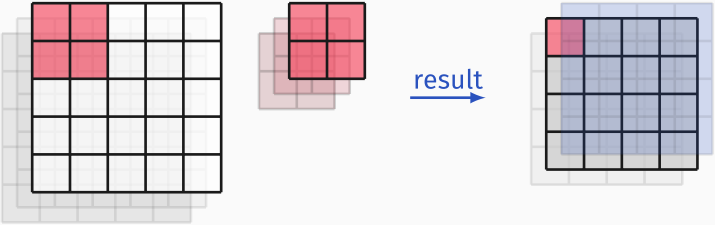
DIRECT CONVOLUTION



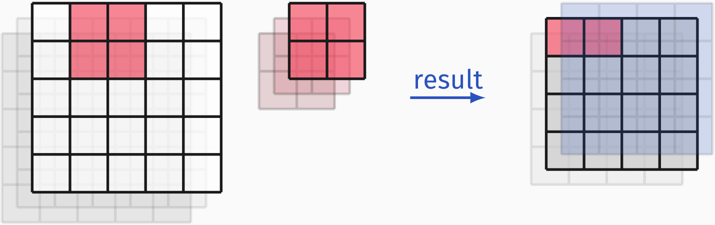
DIRECT CONVOLUTION



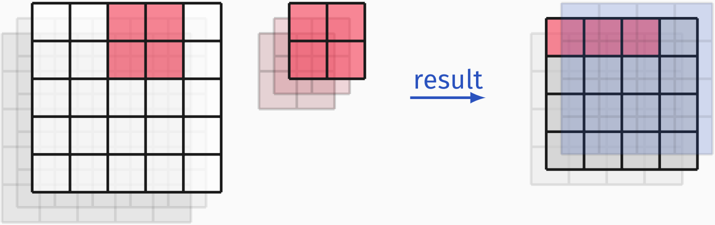
DIRECT CONVOLUTION



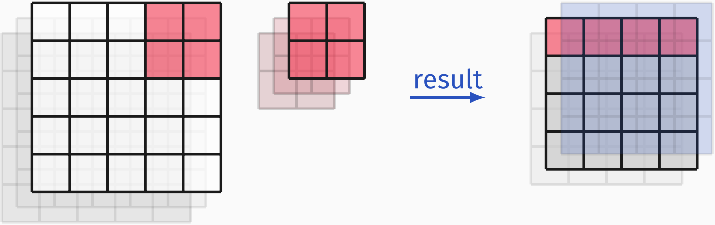
DIRECT CONVOLUTION



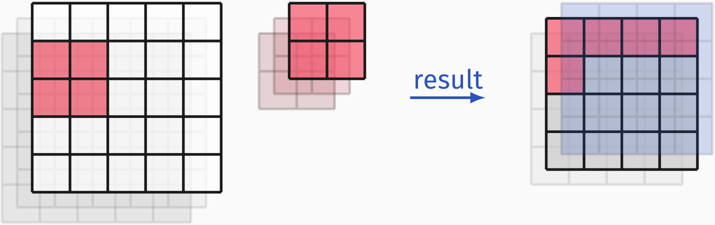
DIRECT CONVOLUTION



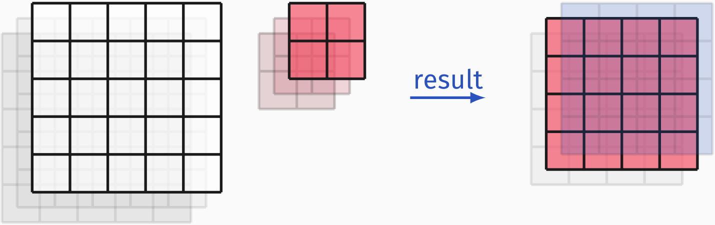
DIRECT CONVOLUTION



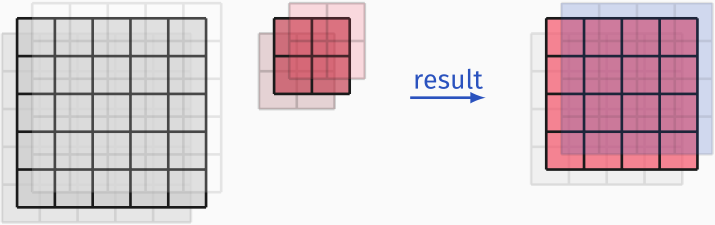
DIRECT CONVOLUTION



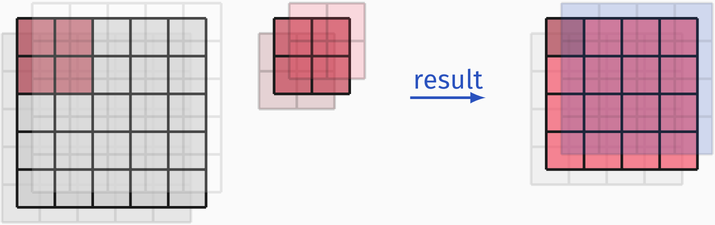
DIRECT CONVOLUTION



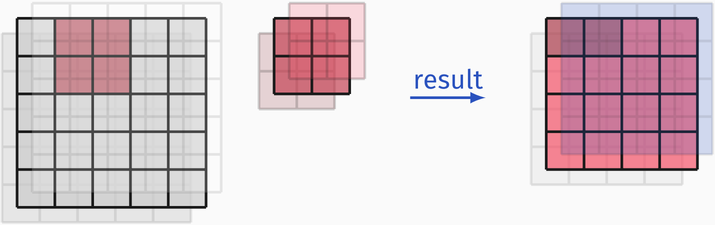
DIRECT CONVOLUTION



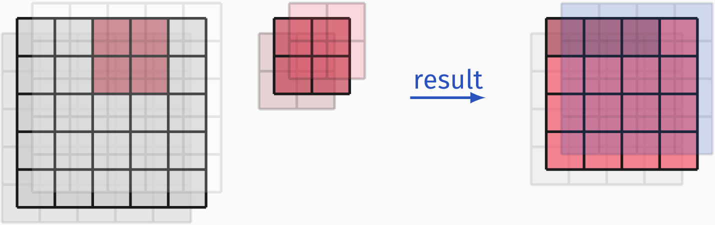
DIRECT CONVOLUTION



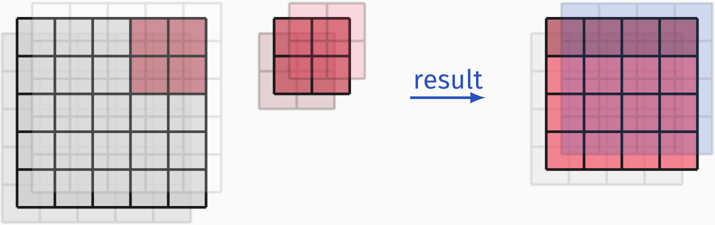
DIRECT CONVOLUTION



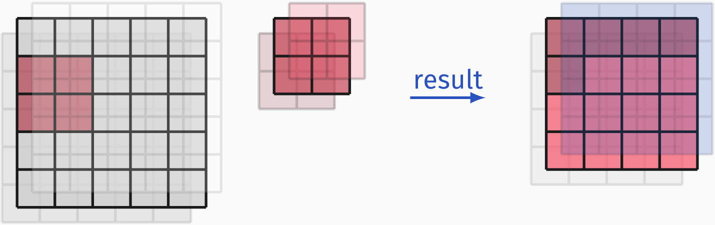
DIRECT CONVOLUTION



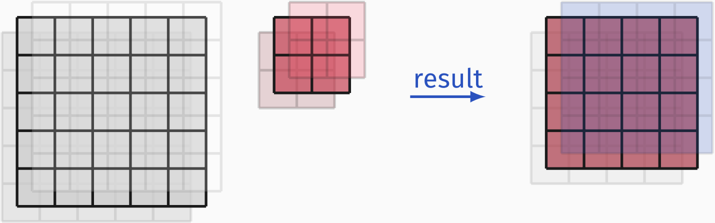
DIRECT CONVOLUTION



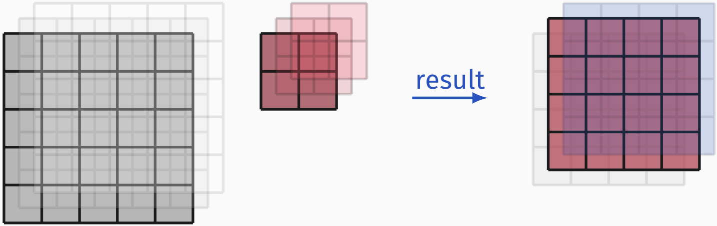
DIRECT CONVOLUTION



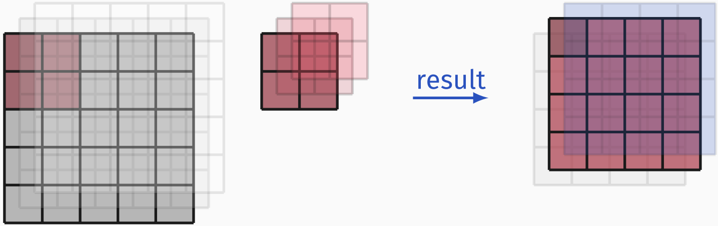
DIRECT CONVOLUTION



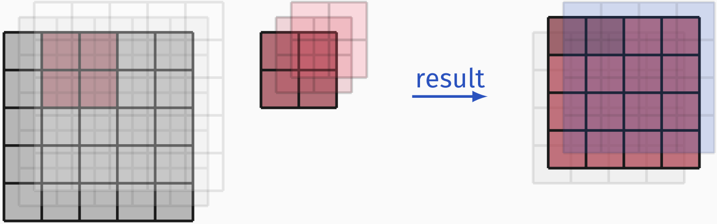
DIRECT CONVOLUTION



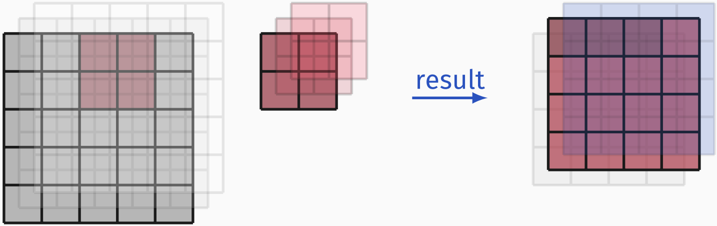
DIRECT CONVOLUTION



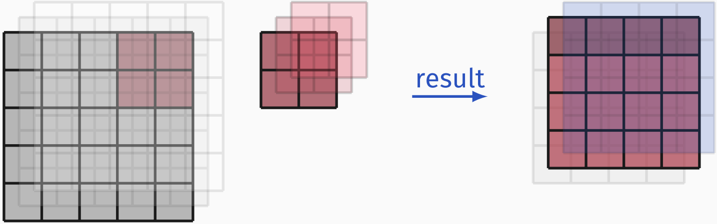
DIRECT CONVOLUTION



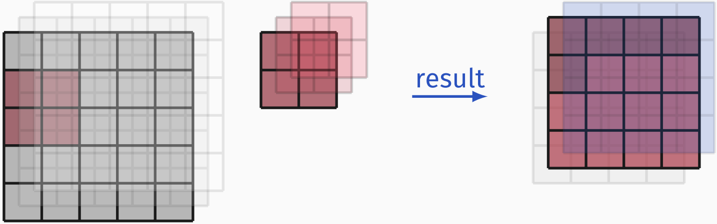
DIRECT CONVOLUTION



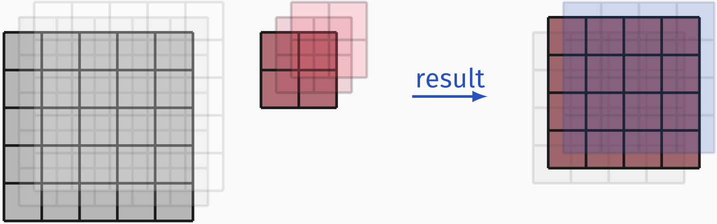
DIRECT CONVOLUTION



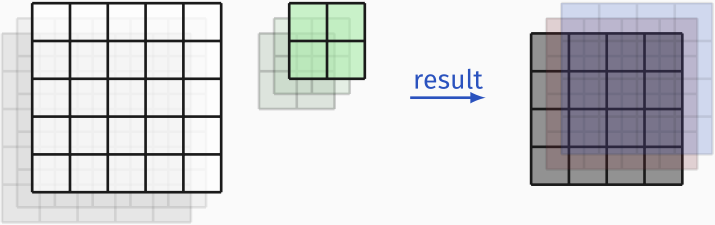
DIRECT CONVOLUTION



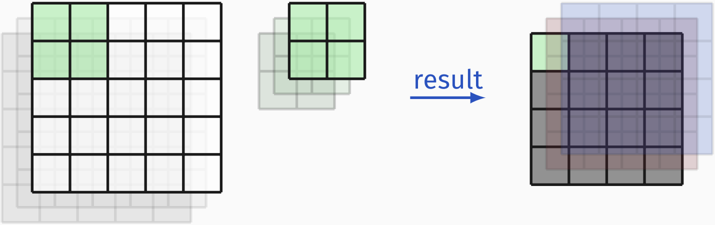
DIRECT CONVOLUTION



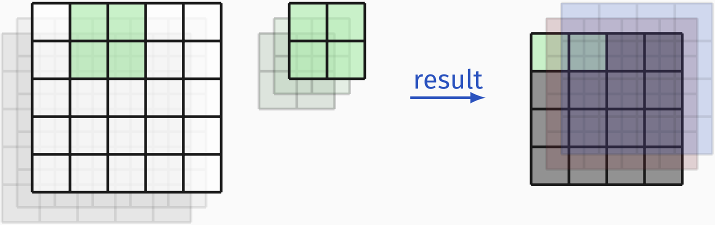
DIRECT CONVOLUTION



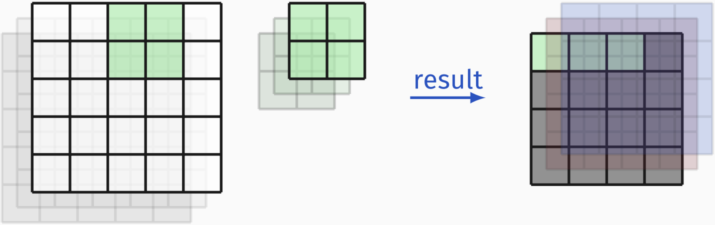
DIRECT CONVOLUTION



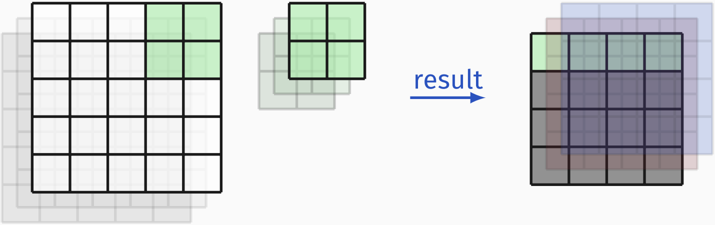
DIRECT CONVOLUTION



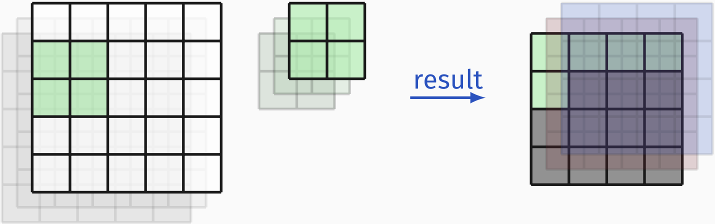
DIRECT CONVOLUTION



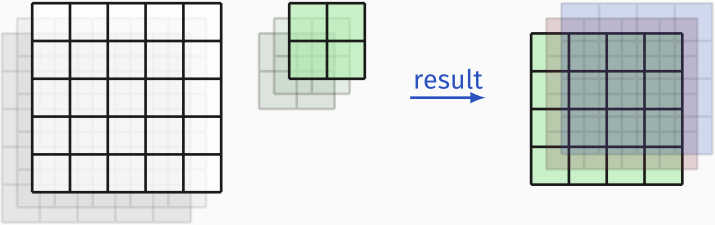
DIRECT CONVOLUTION



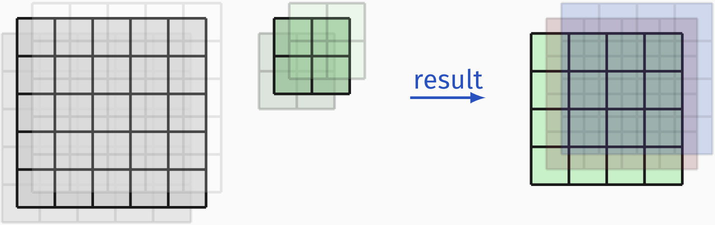
DIRECT CONVOLUTION



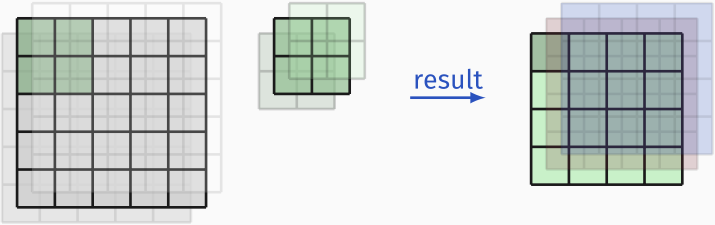
DIRECT CONVOLUTION



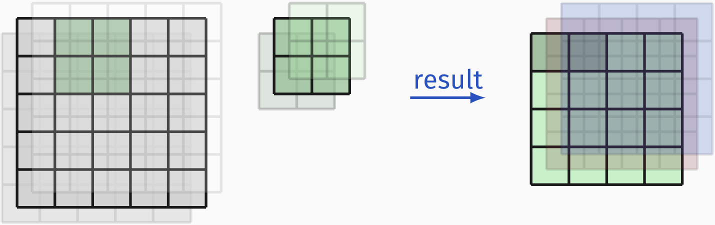
DIRECT CONVOLUTION



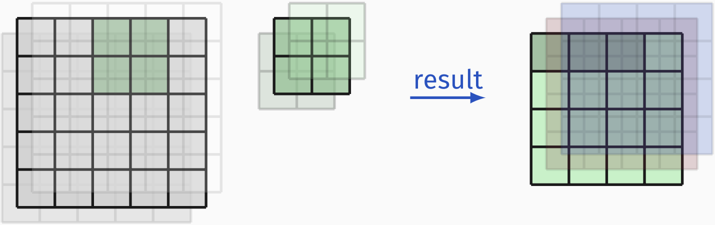
DIRECT CONVOLUTION



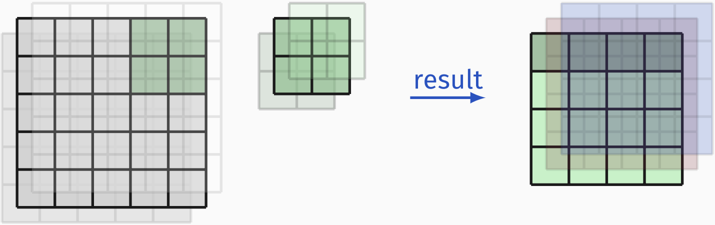
DIRECT CONVOLUTION



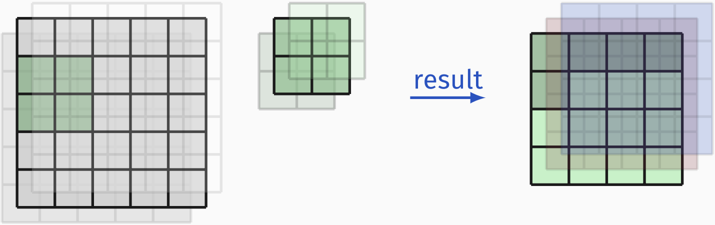
DIRECT CONVOLUTION



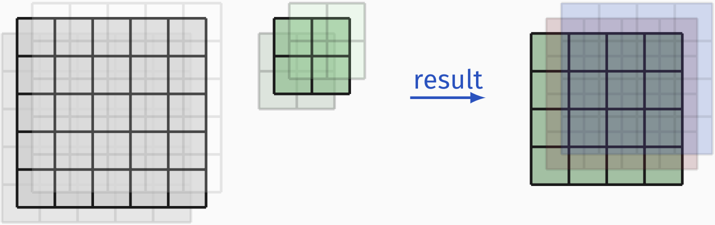
DIRECT CONVOLUTION



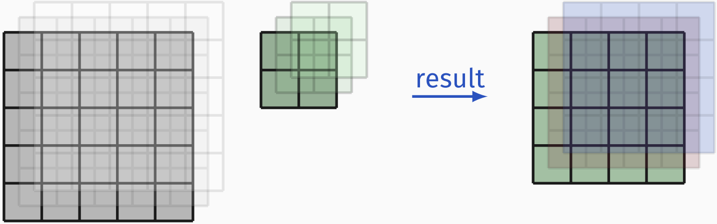
DIRECT CONVOLUTION



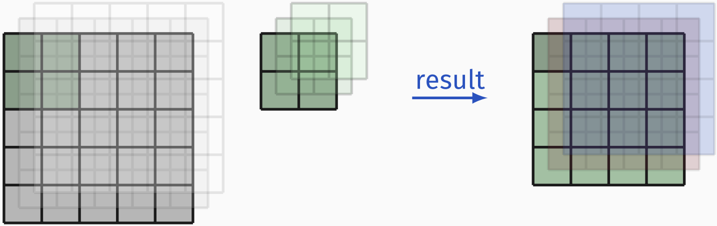
DIRECT CONVOLUTION



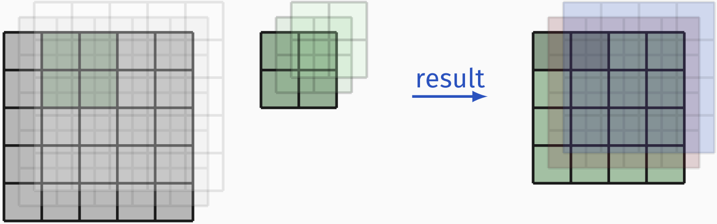
DIRECT CONVOLUTION



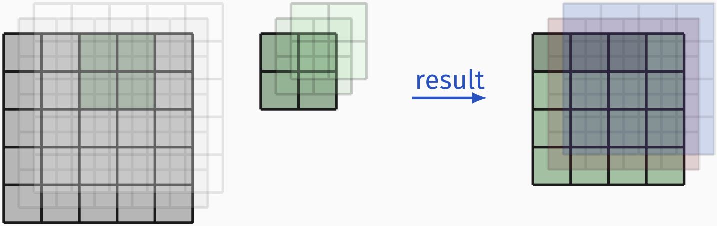
DIRECT CONVOLUTION



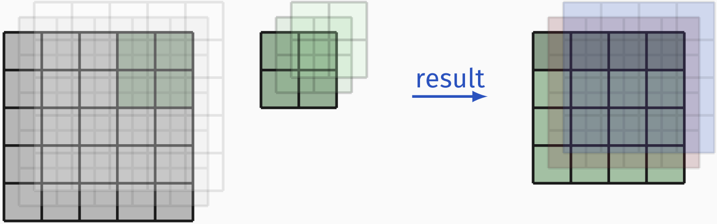
DIRECT CONVOLUTION



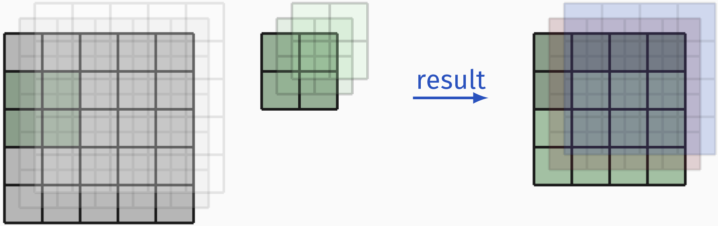
DIRECT CONVOLUTION



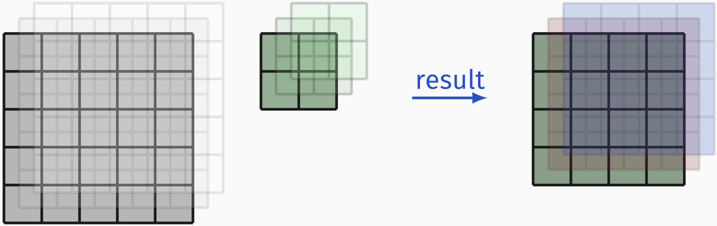
DIRECT CONVOLUTION



DIRECT CONVOLUTION

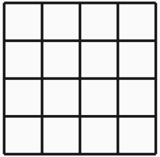


DIRECT CONVOLUTION

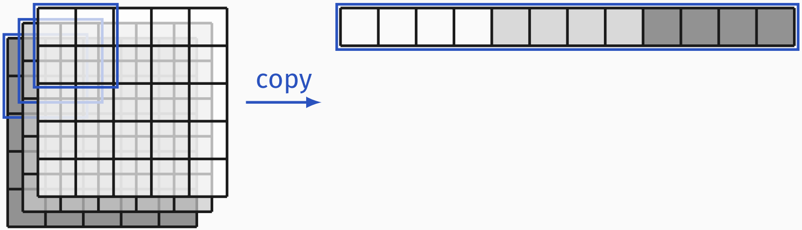


DIRECT CONVOLUTION

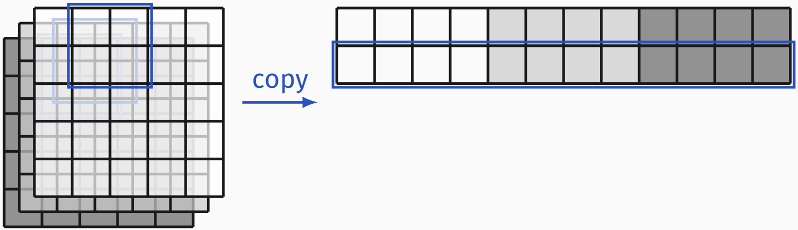
result



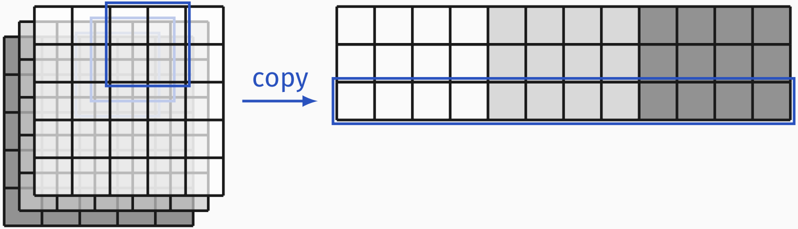
MATRIX MULTIPLICATION BASED CONVOLUTION



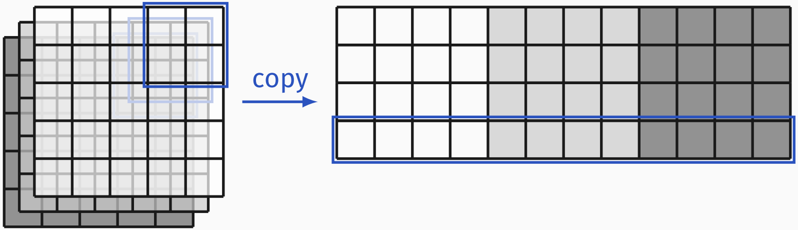
MATRIX MULTIPLICATION BASED CONVOLUTION



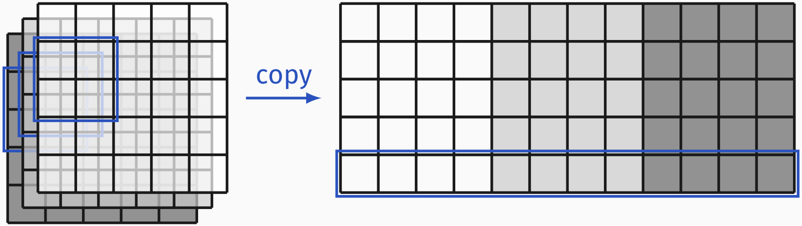
MATRIX MULTIPLICATION BASED CONVOLUTION



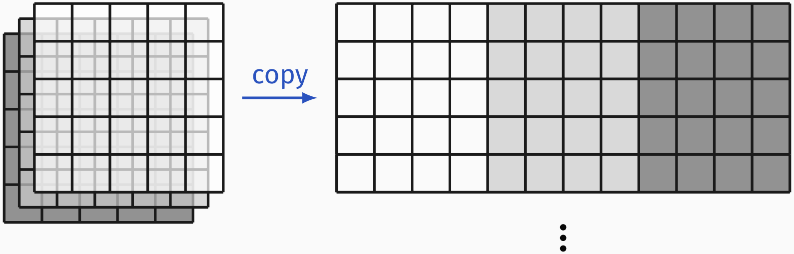
MATRIX MULTIPLICATION BASED CONVOLUTION



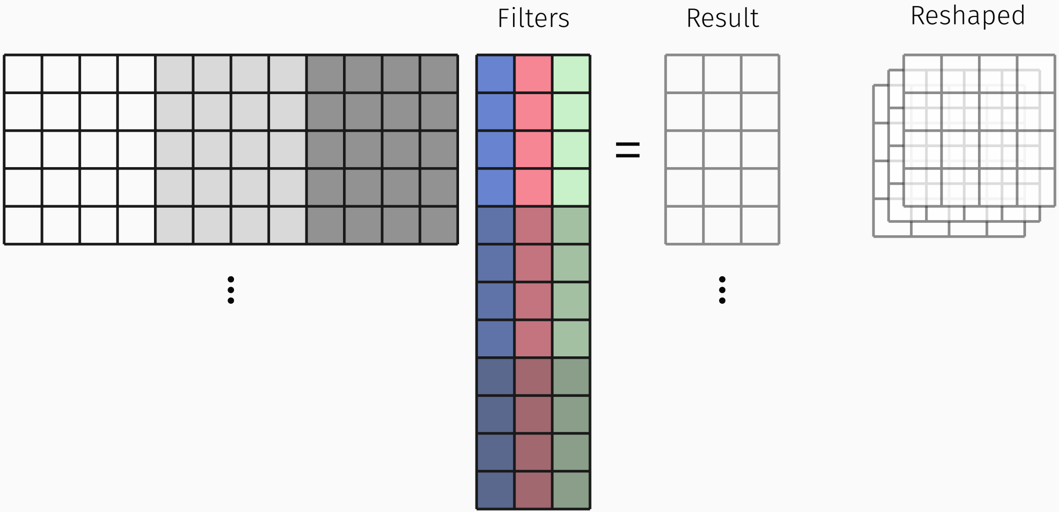
MATRIX MULTIPLICATION BASED CONVOLUTION



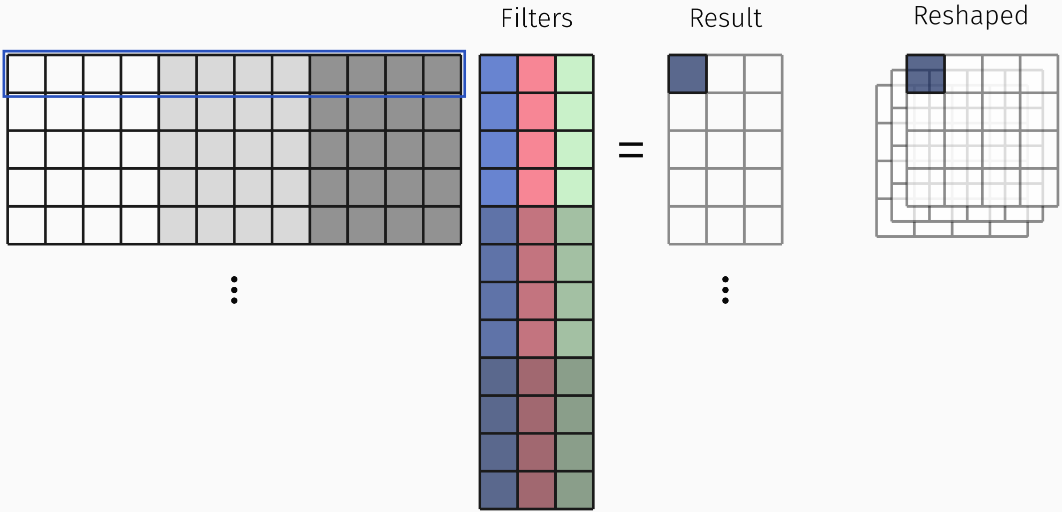
MATRIX MULTIPLICATION BASED CONVOLUTION



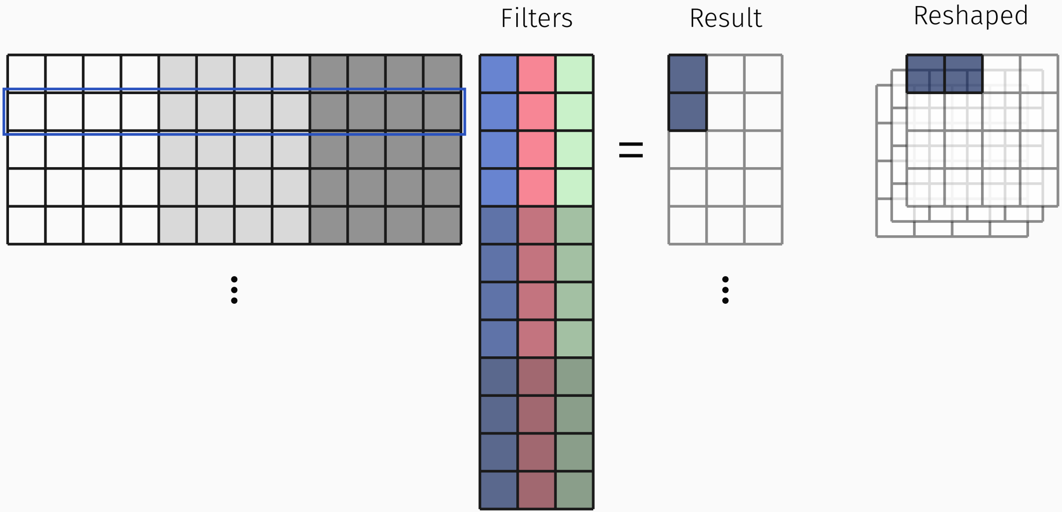
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



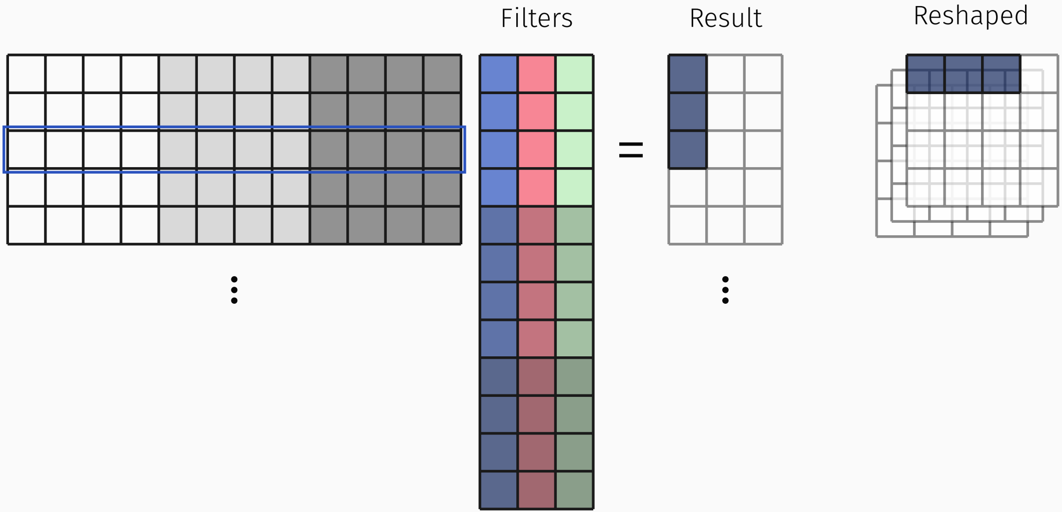
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



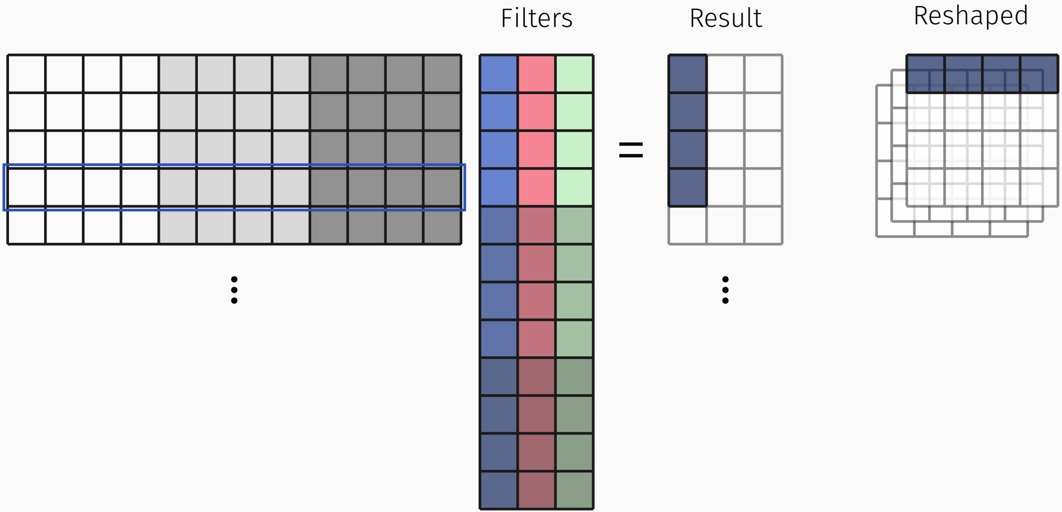
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



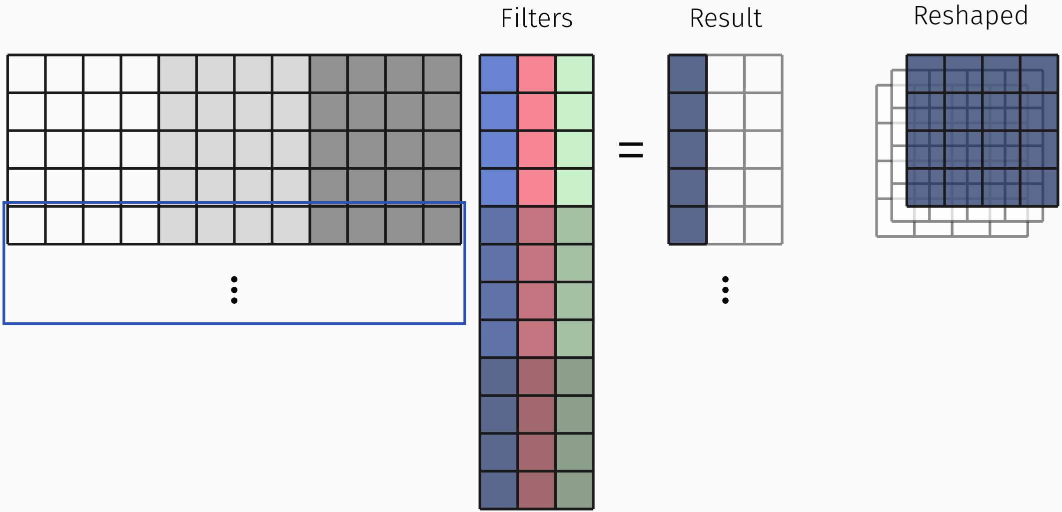
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



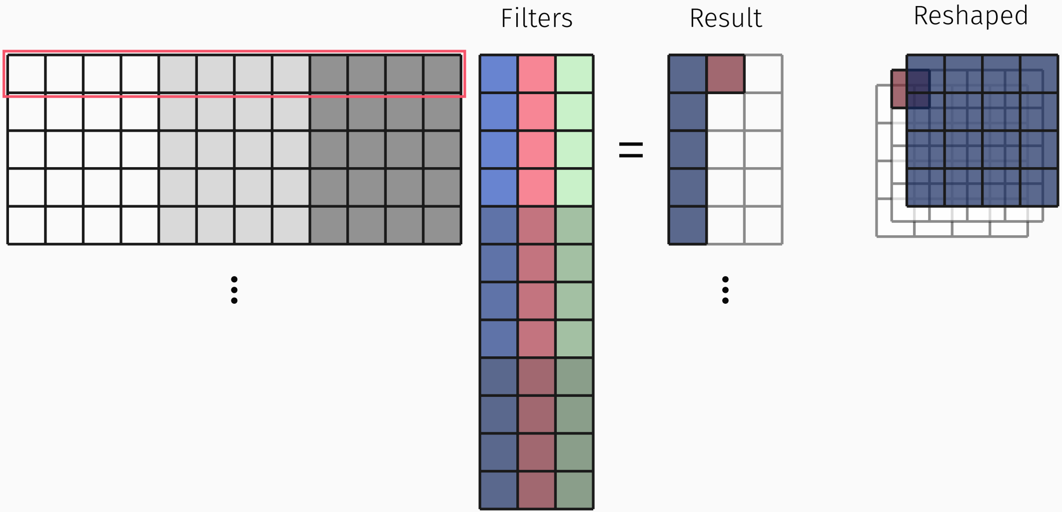
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



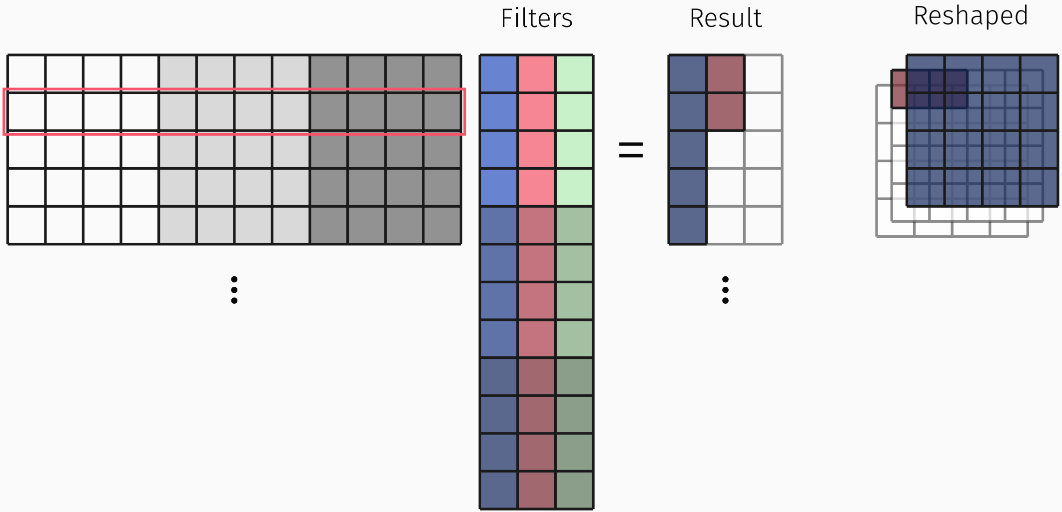
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



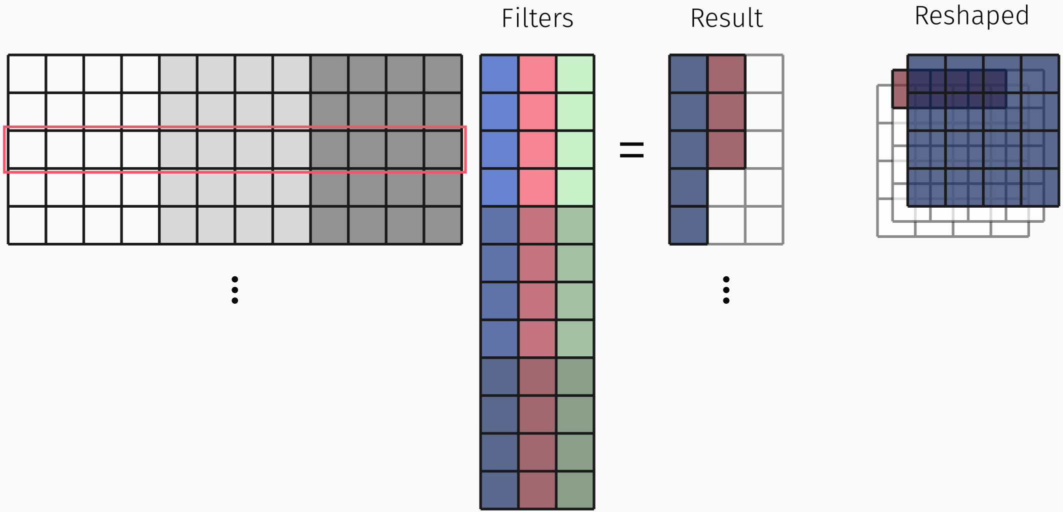
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



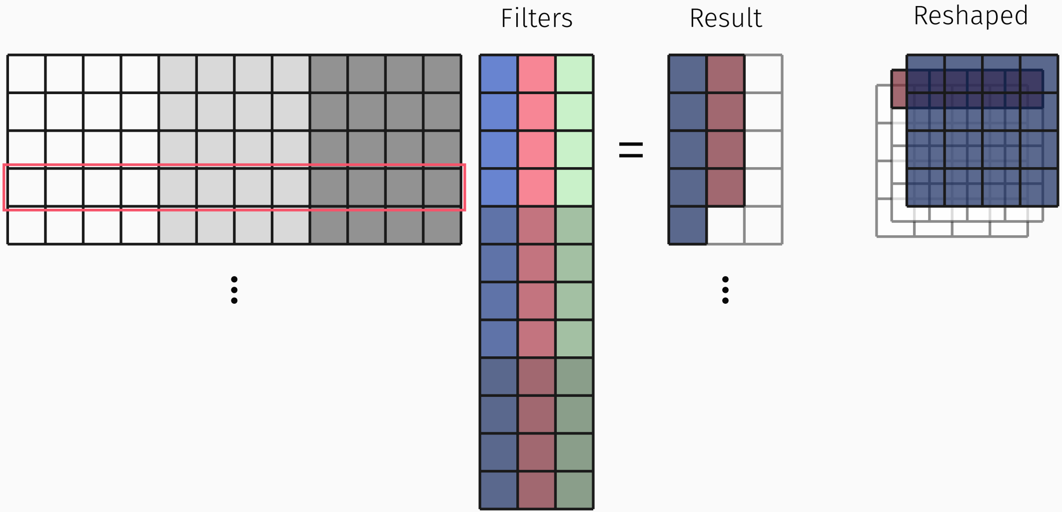
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



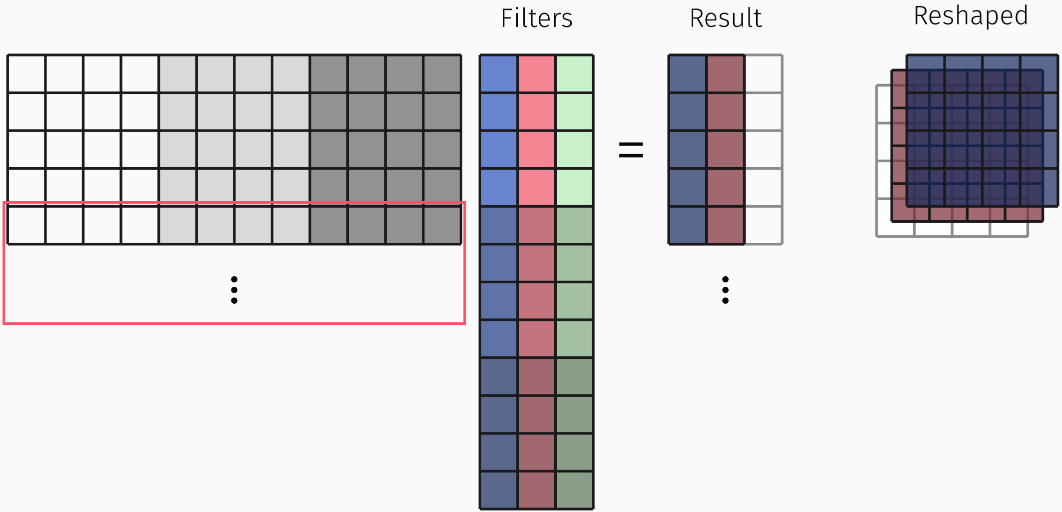
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



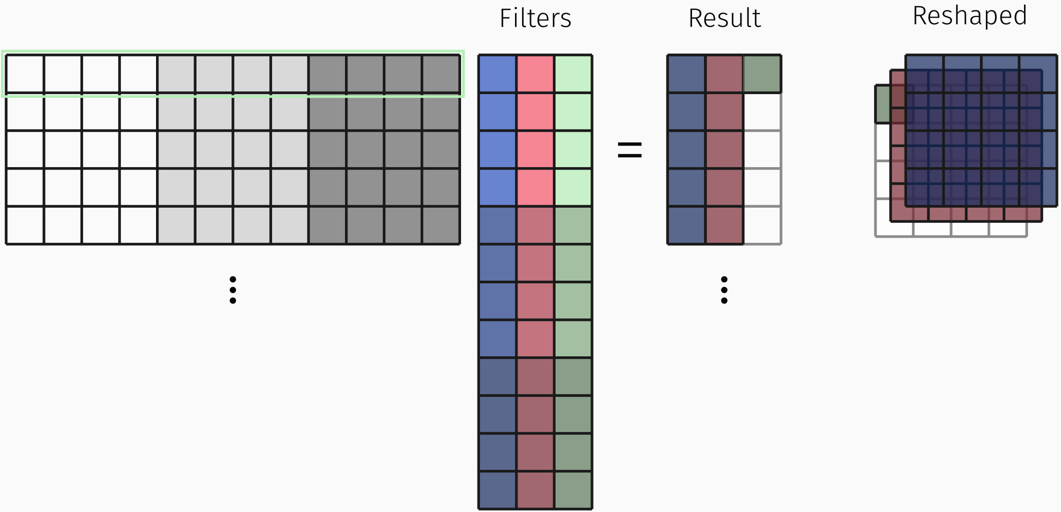
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



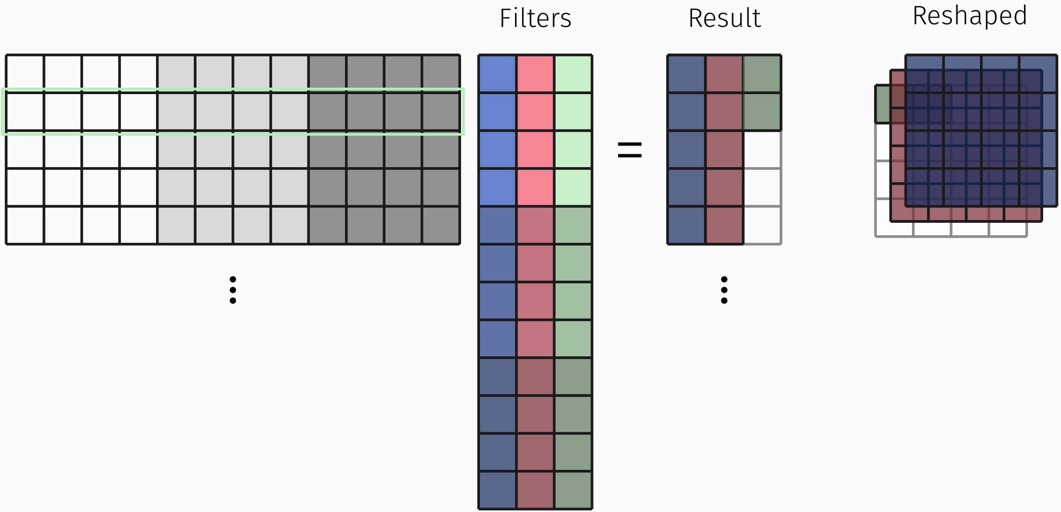
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



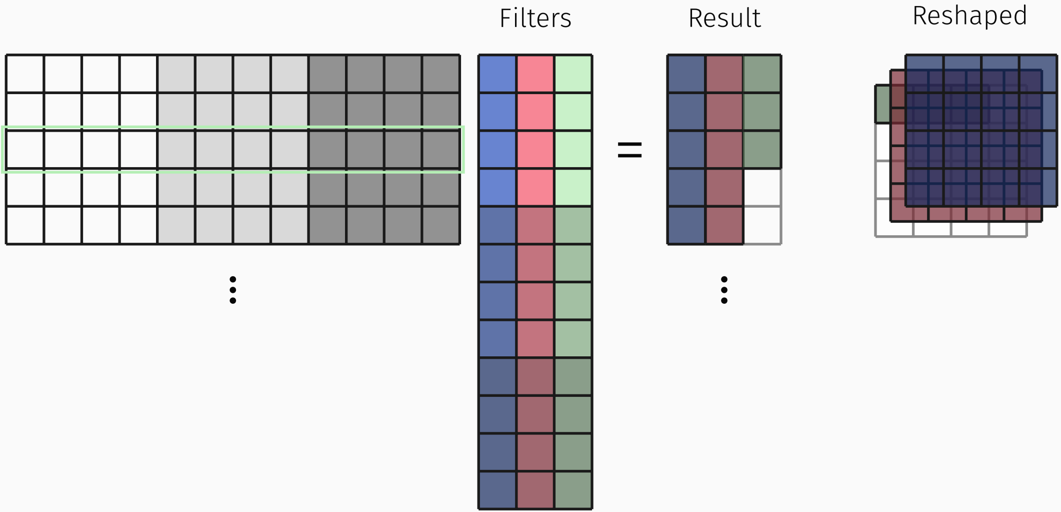
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



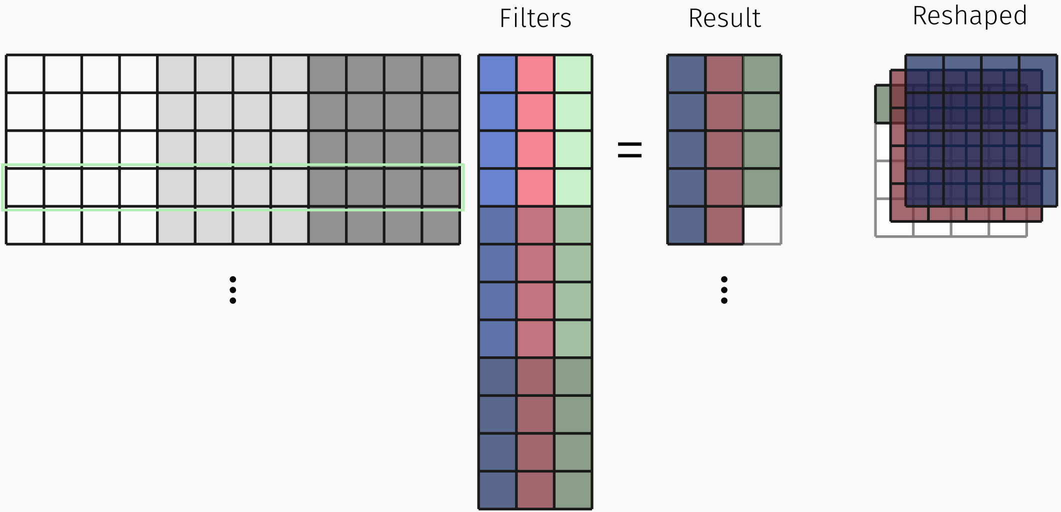
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



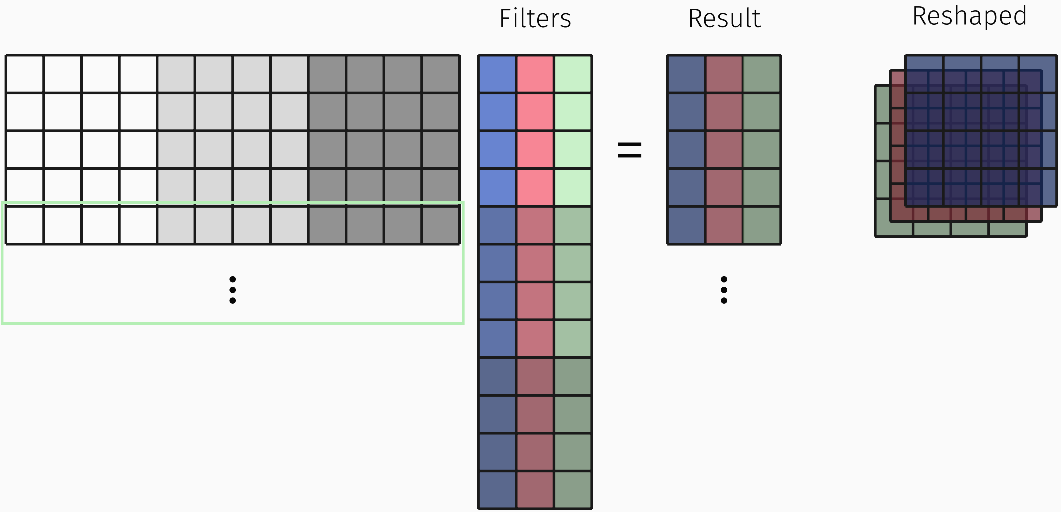
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



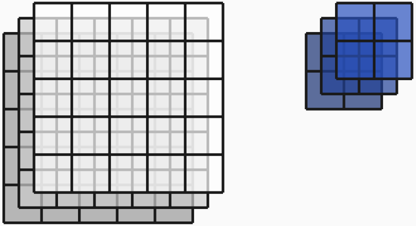
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



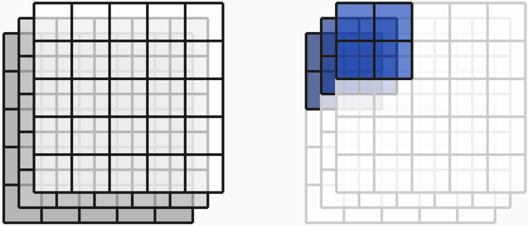
MATRIX MULTIPLICATION BASED CONVOLUTION (CONT'D)



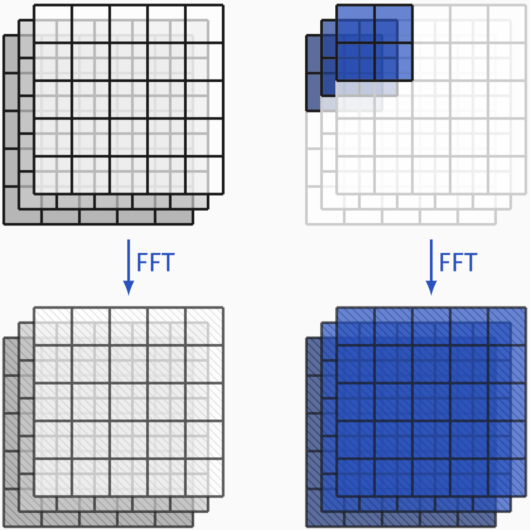
FFT CONVOLUTION



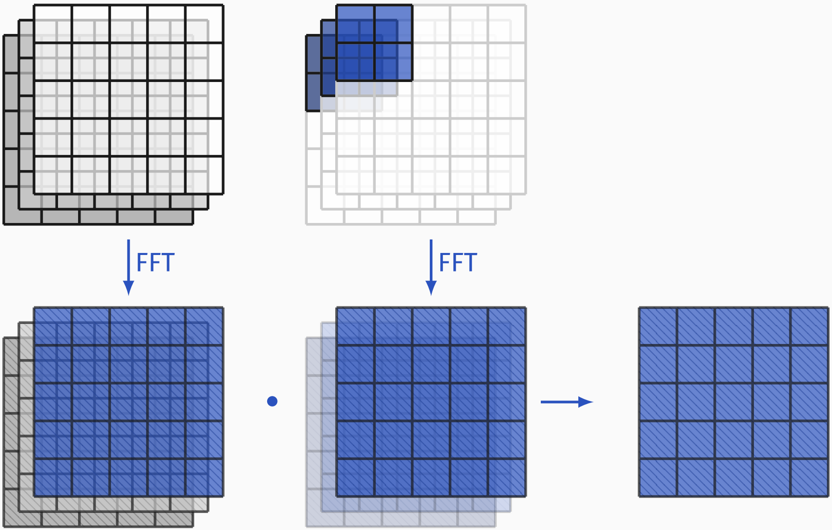
FFT CONVOLUTION



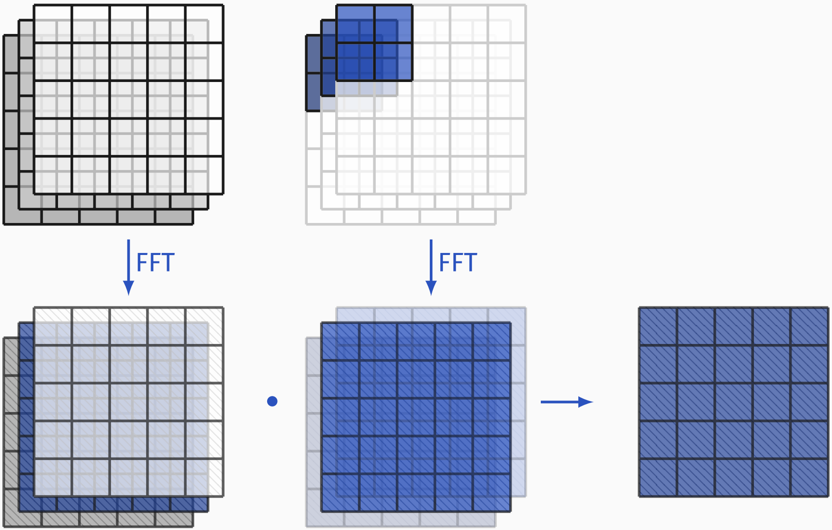
FFT CONVOLUTION



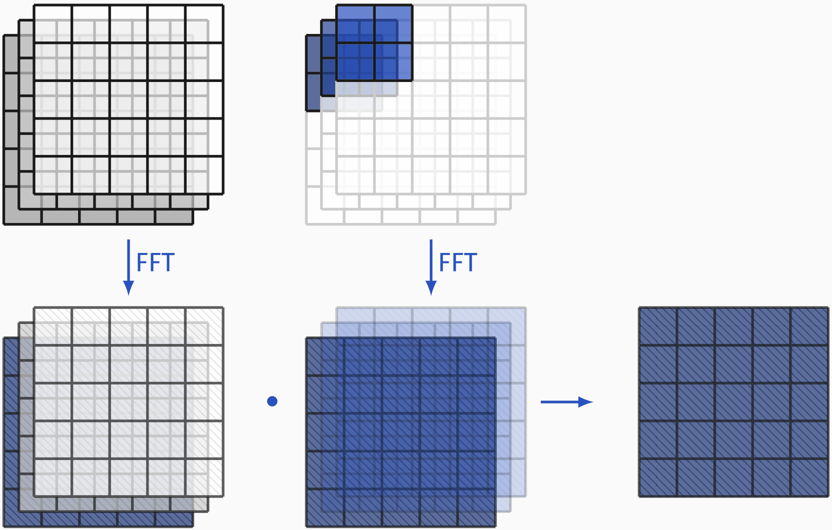
FFT CONVOLUTION



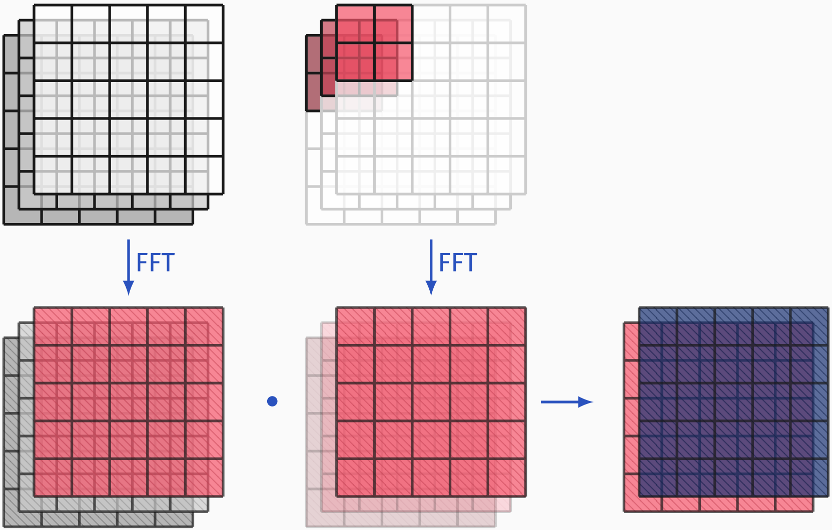
FFT CONVOLUTION



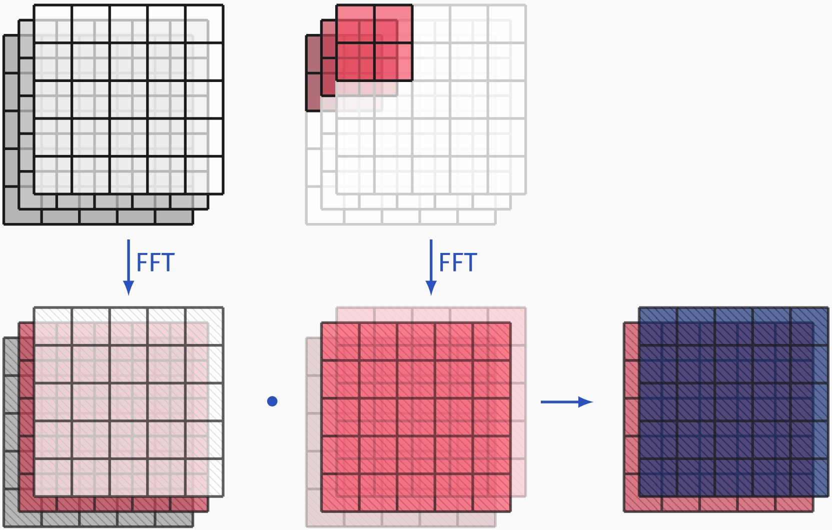
FFT CONVOLUTION



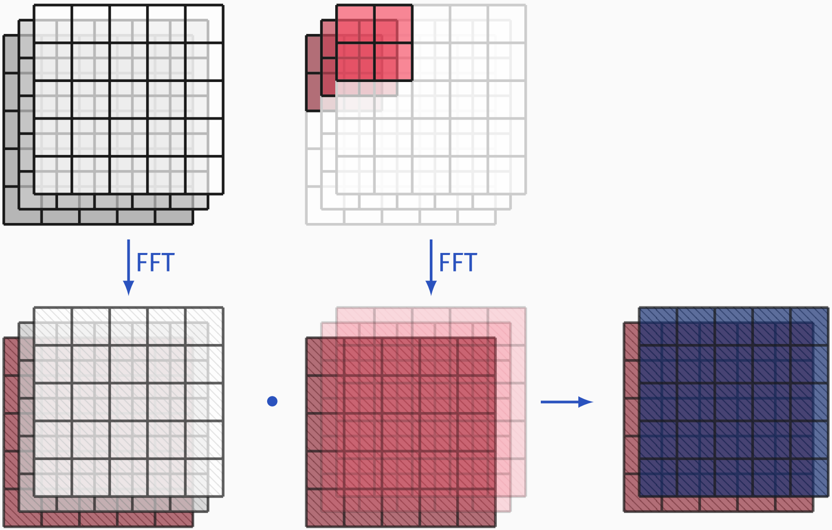
FFT CONVOLUTION



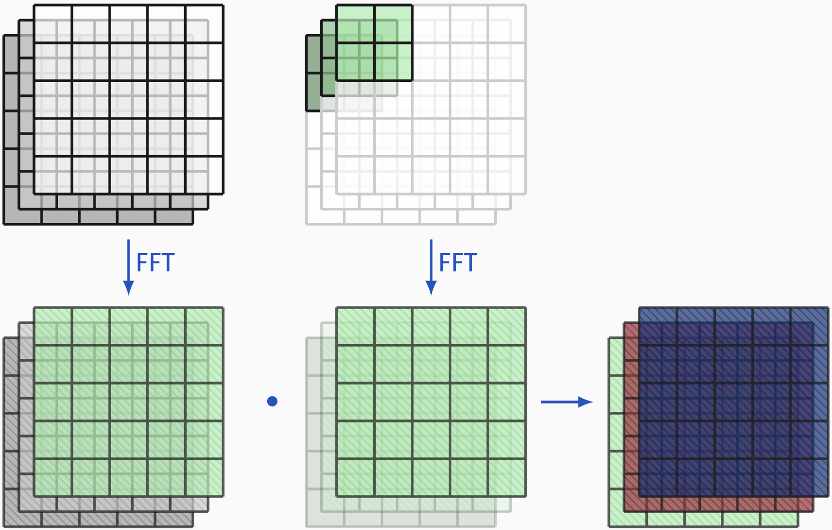
FFT CONVOLUTION



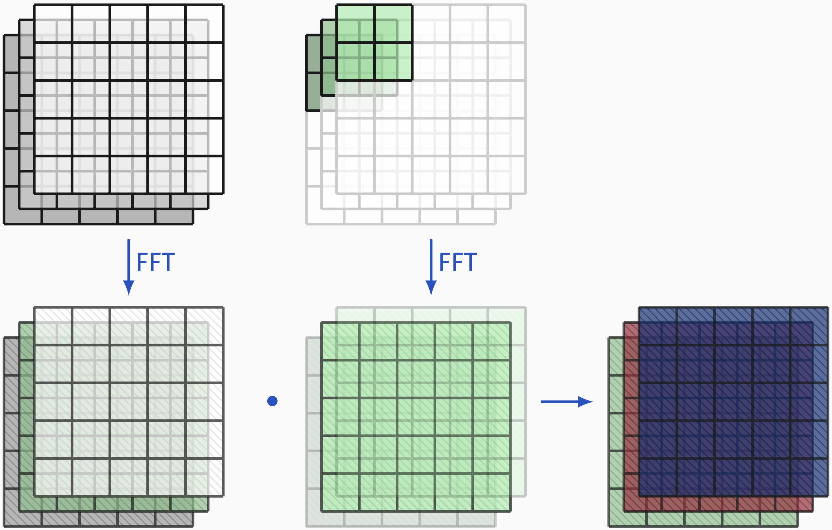
FFT CONVOLUTION



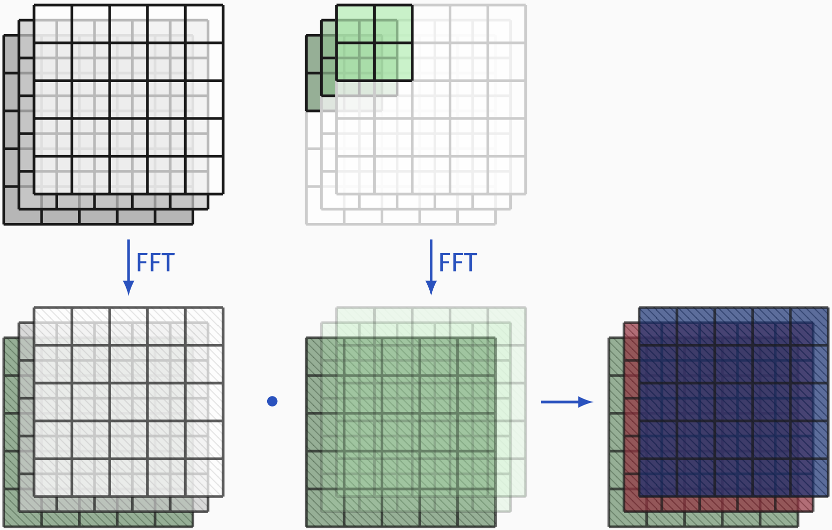
FFT CONVOLUTION



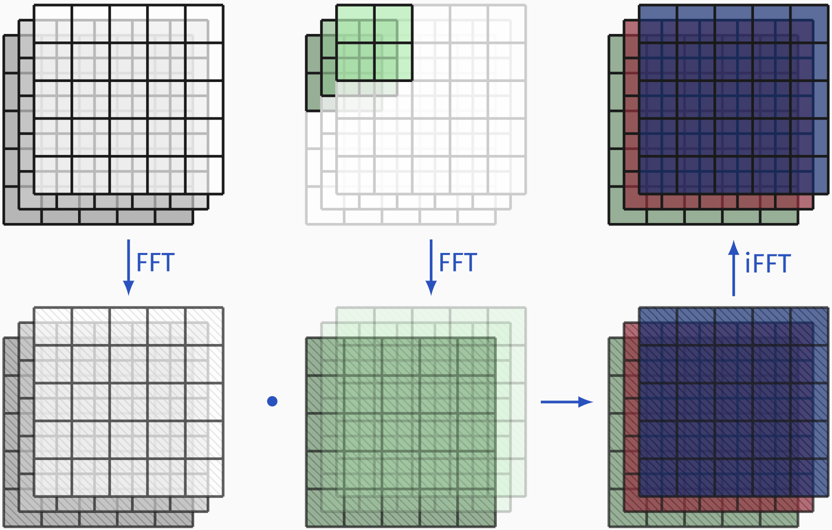
FFT CONVOLUTION



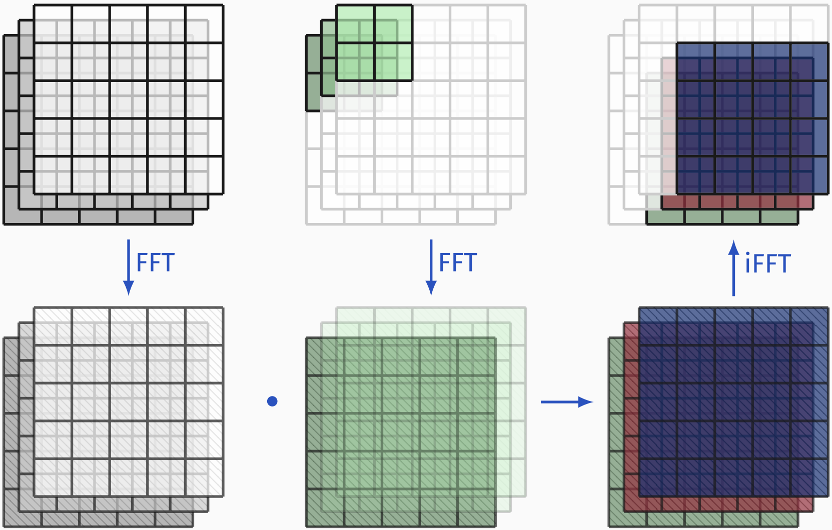
FFT CONVOLUTION



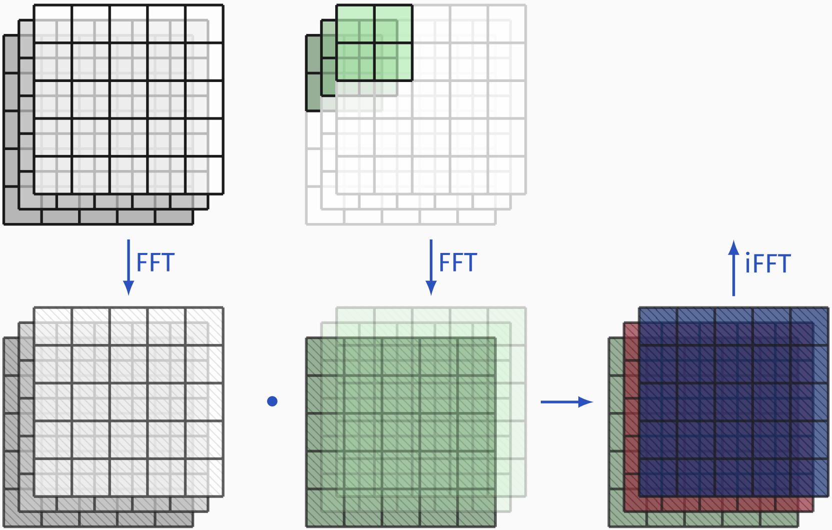
FFT CONVOLUTION



FFT CONVOLUTION



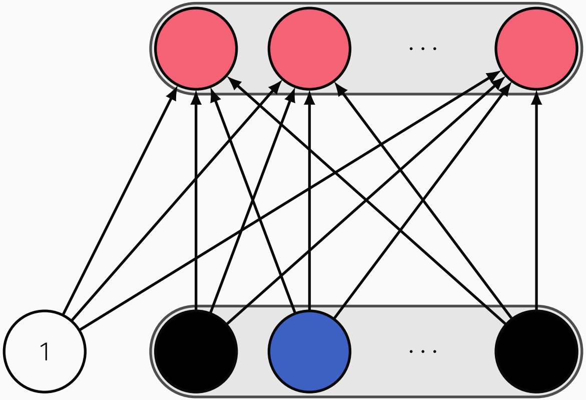
FFT CONVOLUTION



REDUCING OVERFITTING

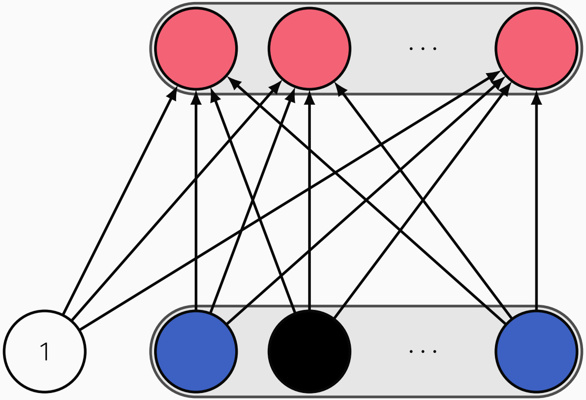
DROPOUT

- Injection of noise
- Randomly set incoming neurons to zero during training with probability $1 - p$
- Rescale neurons/weights by factor $1 - p$ during evaluation



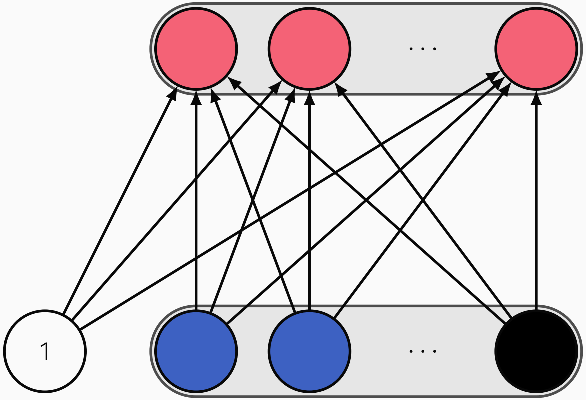
DROPOUT

- Injection of noise
- Randomly set incoming neurons to zero during training with probability $1 - p$
- Rescale neurons/weights by factor $1 - p$ during evaluation



DROPOUT

- Injection of noise
- Randomly set incoming neurons to zero during training with probability $1 - p$
- Rescale neurons/weights by factor $1 - p$ during evaluation



DROPOUT (CONT'D)

- Prevents co-adaptation of neurons (Hinton et al., 2012)
- Linear regression + Dropout = ridge regression with input scale invariant regularization cost (Srivastava, 2013)
- A form of model averaging (with shared weights) (Srivastava et al., 2014)
- Comprehensive overview (Srivastava et al., 2014)

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

Srivastava, N. (2013). Improving neural networks with dropout (Master's Thesis, University of Toronto).

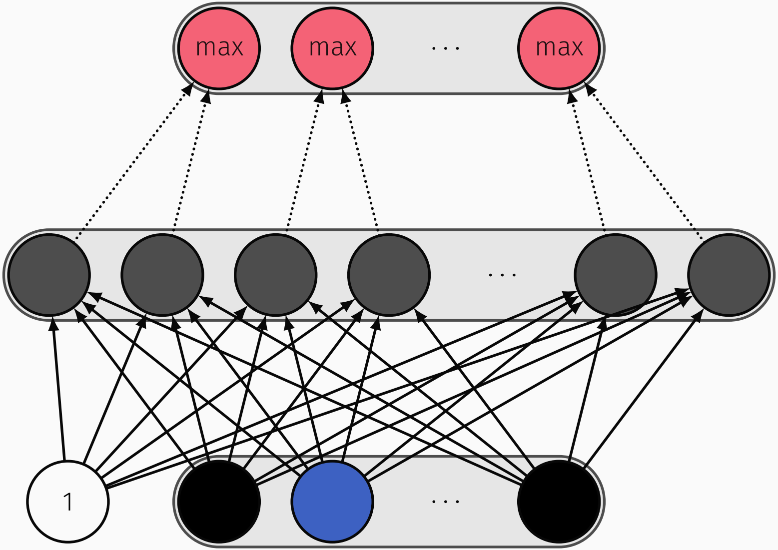
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), pp. 1929-1958.

MAXOUT

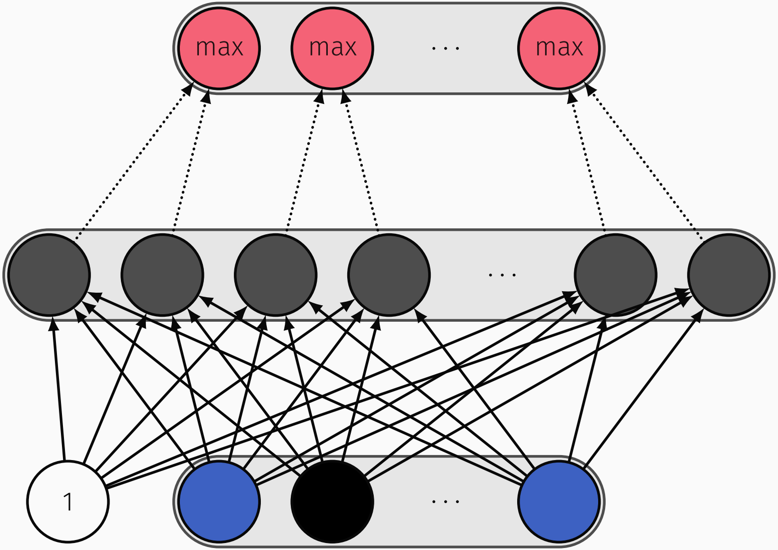
- A learned piecewise linear activation function (Goodfellow et al., 2013)
- Dropout's model averaging approximation is more accurate with Maxout than with tanh
- Gradient flow is better with Maxout than with ReLU

Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout networks. *In* Proceedings of the International Conference on Machine Learning, pp. 1319-1327

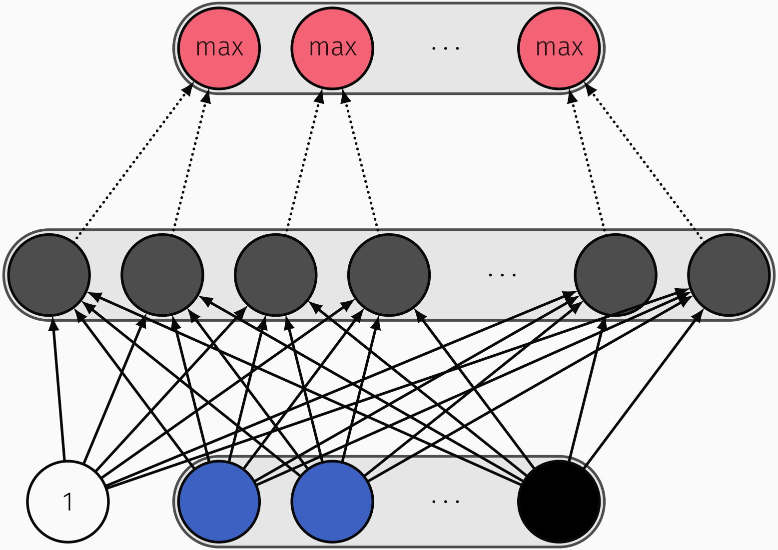
MAXOUT (CONT'D)



MAXOUT (CONT'D)



MAXOUT (CONT'D)



WEIGHT NORM CONSTRAINTS

- Constrain weights vector $\mathbf{w}$: $\|\mathbf{w}\|_2^2 \leq c$ with $c > 0$
- $\mathbf{w}$ is the vector of weights incident to a hidden unit
- Allows to use larger learning rate
- Particularly effective when combined with Dropout (Srivastava et al., 2014)

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), pp. 1929-1958.

DATA AUGMENTATION

- Image translations, i.e. extract smaller "sub-images" (large patches)
- Reflections, i.e. flip images horizontally or vertically (if sensible)
- Intensity perturbation of RGB channels, i.e. "add multiples of the found principal components [of the 3×3 covariance matrix of RGB pixel values], with magnitudes proportional to the corresponding eigenvalues times a random variable drawn from $[\mathcal{N}(0, 0.1)]$." (Krizhevsky et al., 2012)
- Exploitation of symmetries and other invariances

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, pp. 1097-1105.

ARCHITECTURE GUIDELINES

- Increase number of feature maps in deeper layers (e.g. $3 \to 32 \to 64 \to 128$)
- Decrease filter size in deeper layers (e.g. $7 \to 5 \to 3$)
- Roughly keep $k^{(l-1)} \cdot m^{(l-1)} \cdot n^{(l-1)} \approx k^{(l)} \cdot m^{(l)} \cdot n^{(l)}$
- Typically use “valid” or “same” convolution mode
- Typically use strides equal to 1

POOLING

- Max-pooling with non-overlapping regions (typically 2×2)
- Average pooling
- Lp-pooling (Sermanet et al., 2012)
- Stochastic max-pooling (Zeiler & Fergus, 2013)

Sermanet, P., Chintala, S., & LeCun, Y. (2012). Convolutional neural networks applied to house numbers digit classification. *In* International Conference on Pattern Recognition (ICPR), pp. 3288-3291. IEEE.

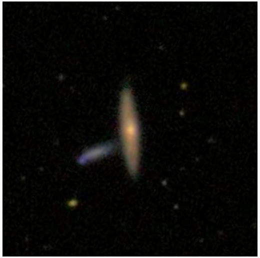
Zeiler, M. D., & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. *In* Proceedings of the International Conference on Representation Learning (ICLR).

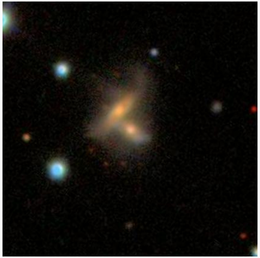
THE GALAXY ZOO CHALLENGE

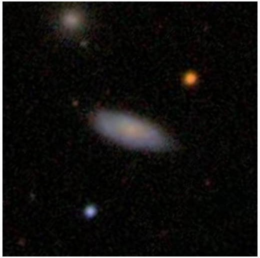
Source: NASA

kaggle

Task: Classify the morphologies of distant galaxies in the universe based on images







THE GALAXY ZOO CHALLENGE (CONT'D)

- 61 578 training images
- 79 975 test images
- Images sized 424×424 pixels
- 37 hierarchical categories
- RMSE between predicted class probabilities and ground truth

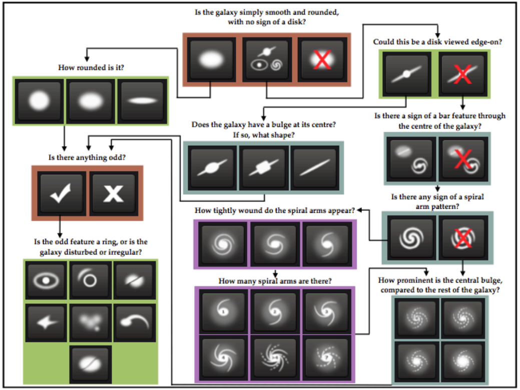
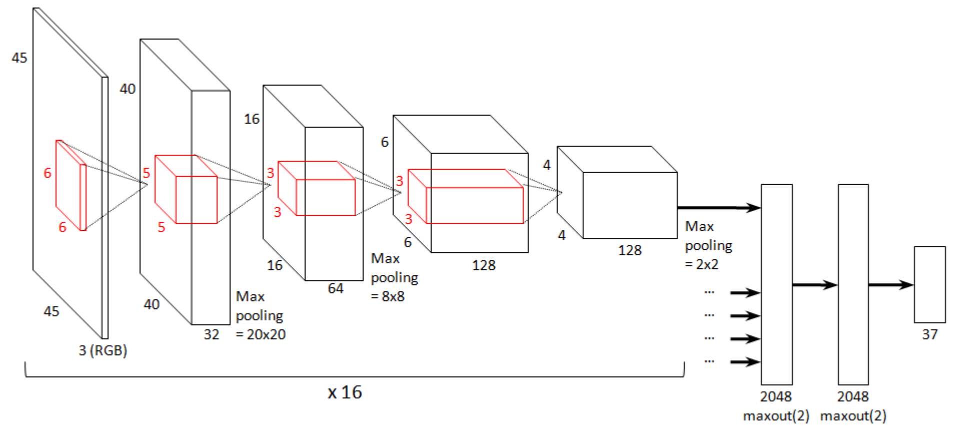


Figure 1. Flowchart of the classification tasks for GZ2, beginning at the top centre. Tasks are colour-coded by their relative depths in the decision tree. Tasks outlined in brown are asked of every galaxy. Tasks outlined in green, blue, and purple are (respectively) one, two or three steps below branching points in the decision tree. Table 2 describes the responses that correspond to the icons in this diagram.

Decision tree of classifications (Willett et al., 2013)

Willett, K. W., Lintott, C. J., Bamford, S. P., Masters, K. L., Simmons, B. D., Casteels, K. R., ... & Thomas, D. (2013). Galaxy Zoo 2: Detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, stt1458.



Best single-model architecture by Sander Dieleman

http://benanne.github.io/2014/04/05/galaxy-zoo.html

Data augmentation: Exploiting spatial invariances

- Images cropped to 207×207
- Galaxy images are rotation invariant $\Rightarrow$ random rotations
- Galaxy images are (to some extent) translation invariant $\Rightarrow$ random shifts by ± 4 pixels (relative to original size)
- Galaxy images are (to some extent) zoom invariant $\Rightarrow$ random (log-uniform) scaling with factors in the range $[\frac{1}{1.3}, 1.3]$
- Galaxy images are flip invariant $\Rightarrow$ random flipping
- Finally, downsampled to 69×69

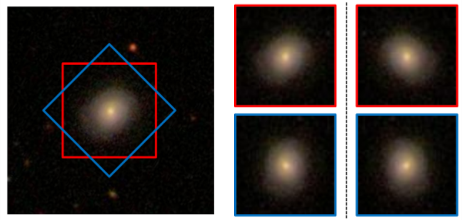
Data augmentation: Color perturbation

- Intensities of RGB channels perturbed
- Multiples of the first principal component (magnitude proportional to the corresponding eigenvalue times a random variable drawn from $\mathcal{N}(0, 0.5)$ added
- Adapted from Krizhevsky et al. (2012)

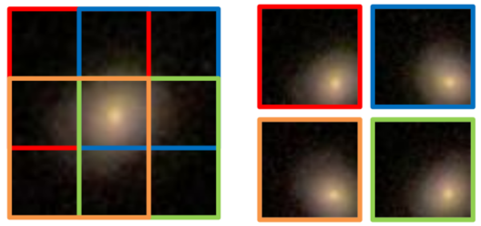
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, pp. 1097-1105.

Network architecture: Exploiting rotation invariance to increase parameter sharing

- Regular crop + 45° rotated crop, both flipped
- Four partially overlapping 45×45 parts

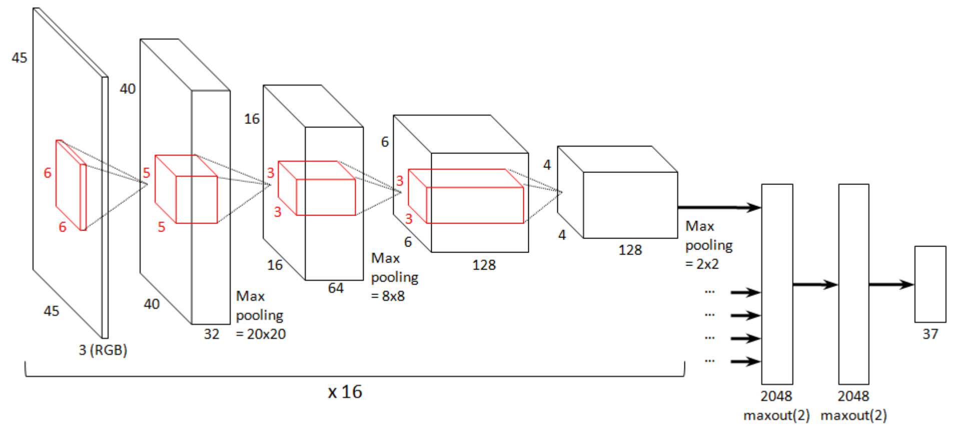


Crop + 45° rotated crop + flip



Partially overlapping parts

THE GALAXY ZOO CHALLENGE (CONT'D)



Best single-model architecture by Sander Dieleman

Incorporating output constraints

- Flat output probabilities: $z'_i = \frac{\max(0, z_i)}{\epsilon + \sum_k \max(0, z_k)}$ for each node in the tree
- Lower level node probabilities are obtained by multiplying the probabilities of the path in the tree

QUESTIONS?