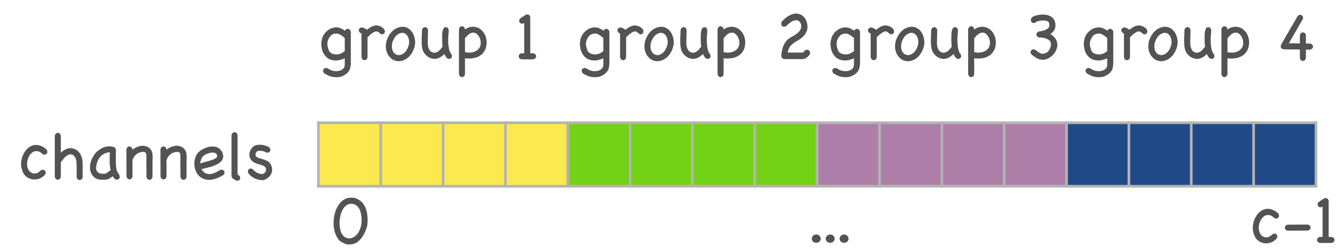
Group normalization and local response normalization

Group normalization



$$\mathbf{Z} \in \mathbb{R}^{B \times W \times H \times C}$$

- Normalize groups of G channels together

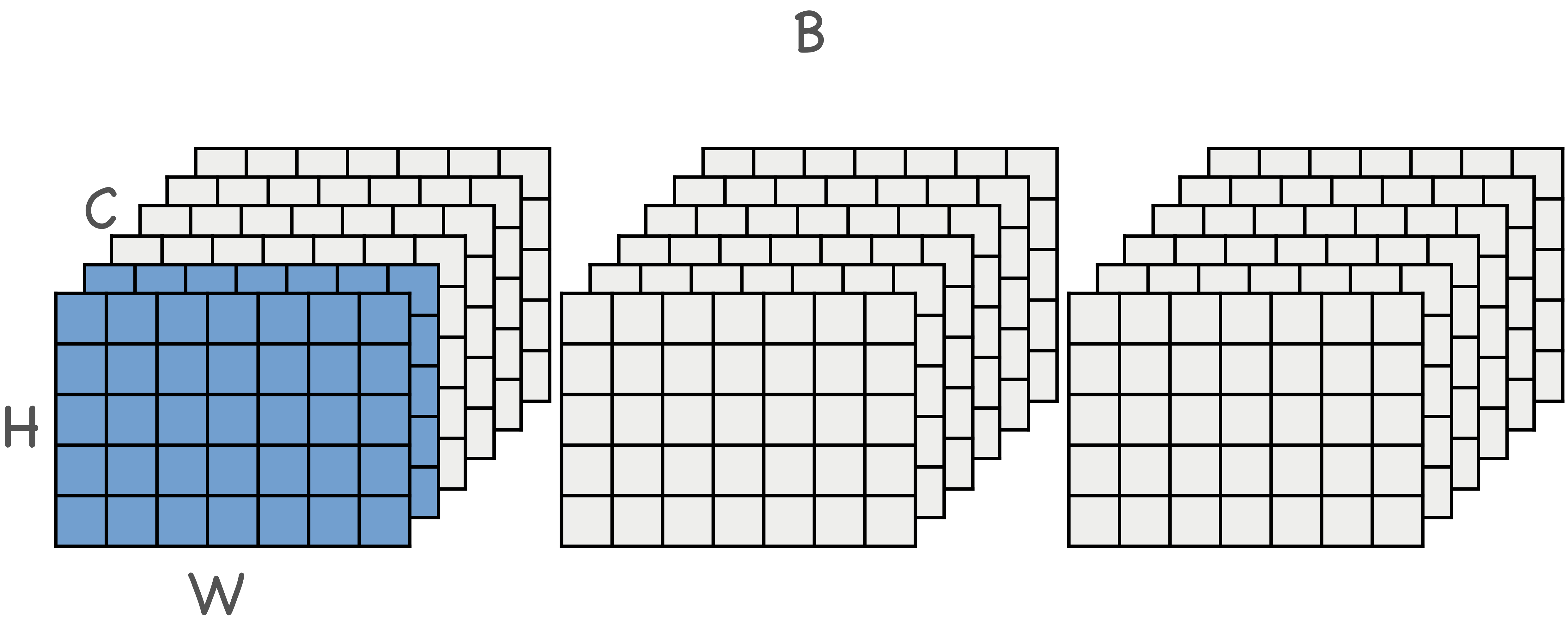
$$\frac{\mathbf{Z}_{k,x,y,c} - \mu_{kg}}{\sigma_{kg}} \qquad g = \lfloor c/G \rfloor$$

$$\mu_{kg} = \frac{1}{WHG} \sum_{c=gG}^{(g+1)G-1} \sum_{x,y} \mathbf{Z}_{k,x,y,c}$$

$$\sigma_{kg}^2 = \frac{1}{WHG} \sum_{c=gG}^{(g+1)G-1} \sum_{x,y} (\mathbf{Z}_{k,x,y,c} - \mu_{kg})^2$$

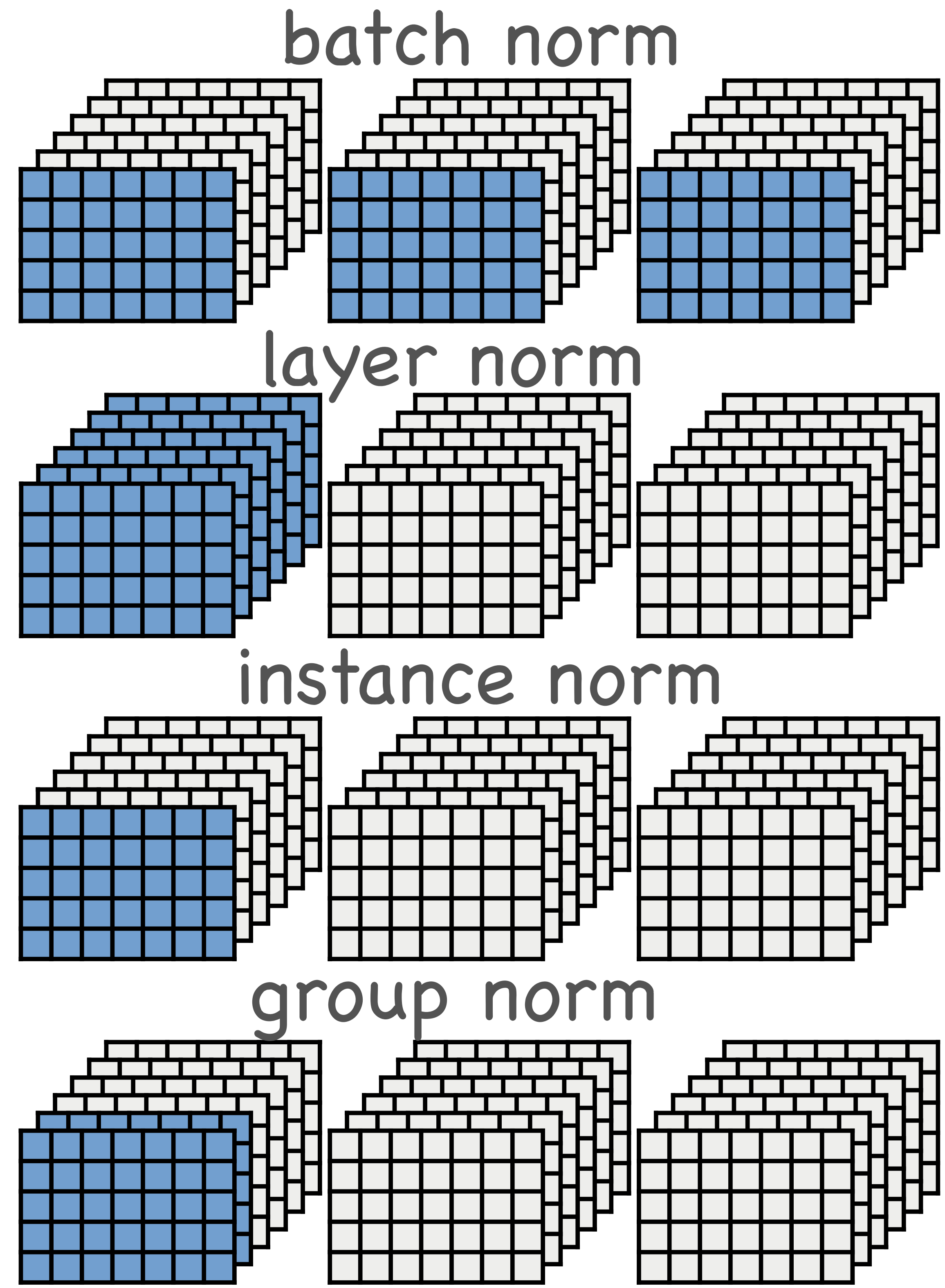
Yuxin Wu, and Kaiming He. "Group normalization." ECCV. 2018.

What does group normalization do?

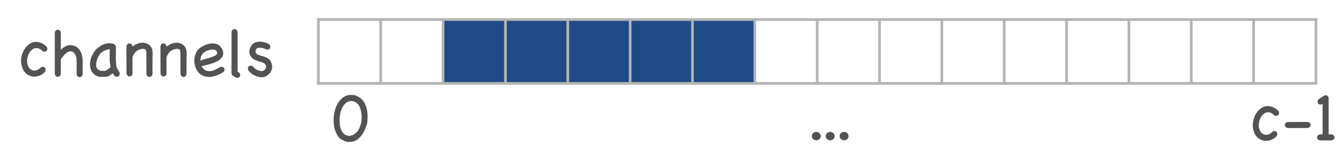


Comparison to other norms

- More stable statistics than instance norm
  - G=C
- Not all channels tied as in layer norm
  - G=1



Local response normalization



- "Generalization" of group norm
- Parameters α and β

$$\mathbf{Z} \in \mathbb{R}^{B \times W \times H \times C}$$

$$\mathbf{Z}_{k,x,y,c} \left(\gamma + \frac{\alpha}{n} \sum_{c'=c-\frac{n}{2}}^{c+\frac{n}{2}} \mathbf{Z}^2_{k,x,y,c} \right)^{-\beta}$$

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." NIPS 2012

Differences between LRN and GN

- Group norm
    - Normalize over all spatial locations
    - Subtract mean
    - Scale and bias transformation
- Local response normalization
    - More flexible parametrization

$$\mathbf{Z} \in \mathbb{R}^{B\times W\times H\times C}$$

$$\mathbf{Z}_{k,x,y,c}\left(\gamma + \frac{\alpha}{n}\sum_{c'=c-\frac{n}{2}}^{c+\frac{n}{2}} \mathbf{Z}^2_{k,x,y,c}\right)^{-\beta}$$