Quelques outils classiques

Batch normalization
problème : *internal covariate shift*

Wu, Y. and He, K., 2018. Group normalization. arXiv preprint arXiv: 1803.08494.

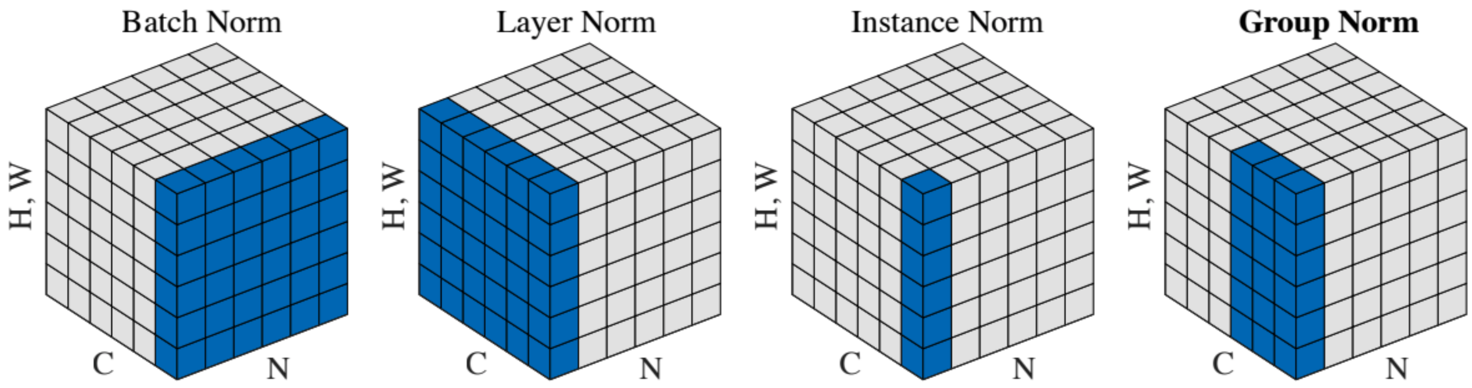


Figure 2. **Normalization methods**. Each subplot shows a feature map tensor, with N as the batch axis, C as the channel axis, and (H, W) as the spatial axes. The pixels in blue are normalized by the same mean and variance, computed by aggregating the values of these pixels.

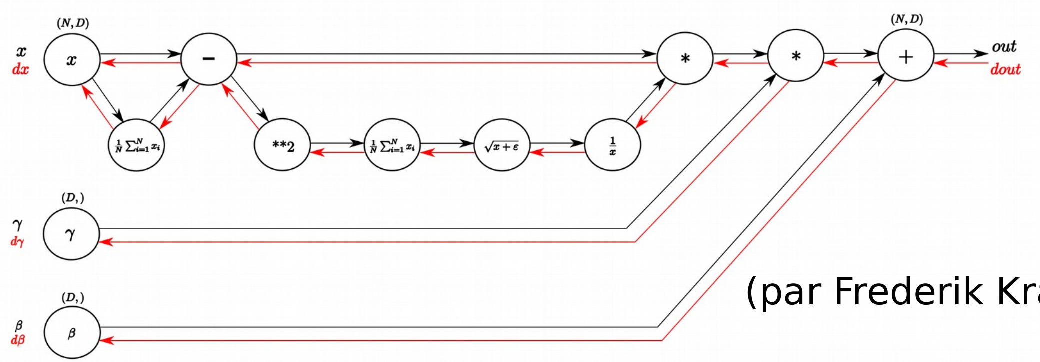
Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_\mathcal{B} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_\mathcal{B}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_\mathcal{B})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_\mathcal{B}}{\sqrt{\sigma_\mathcal{B}^2 + \epsilon}} \qquad \text{// normalize}$$

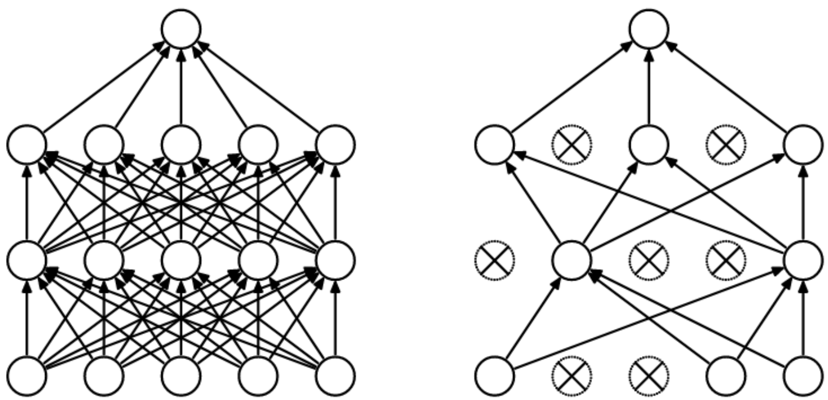
$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$



(par Frederik Kratzert)

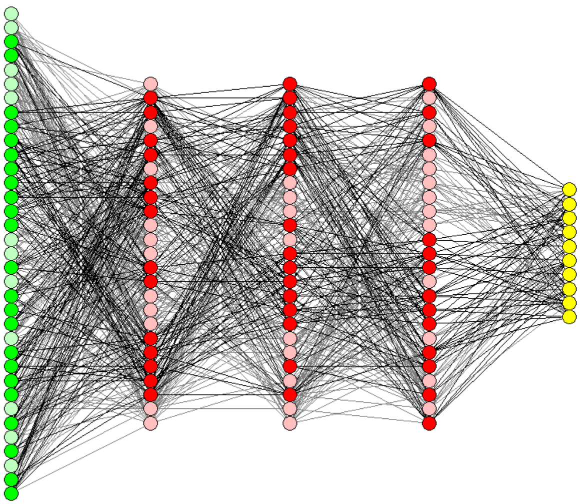
Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift
Sergey Ioffe, Christian Szegedy - https://arxiv.org/abs/1502.03167

Dropouts
réduit la co-adaptation entre les neurones



(a) Standard Neural Net (b) After applying dropout.

Figure 1: Dropout Neural Net Model. **Left**: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

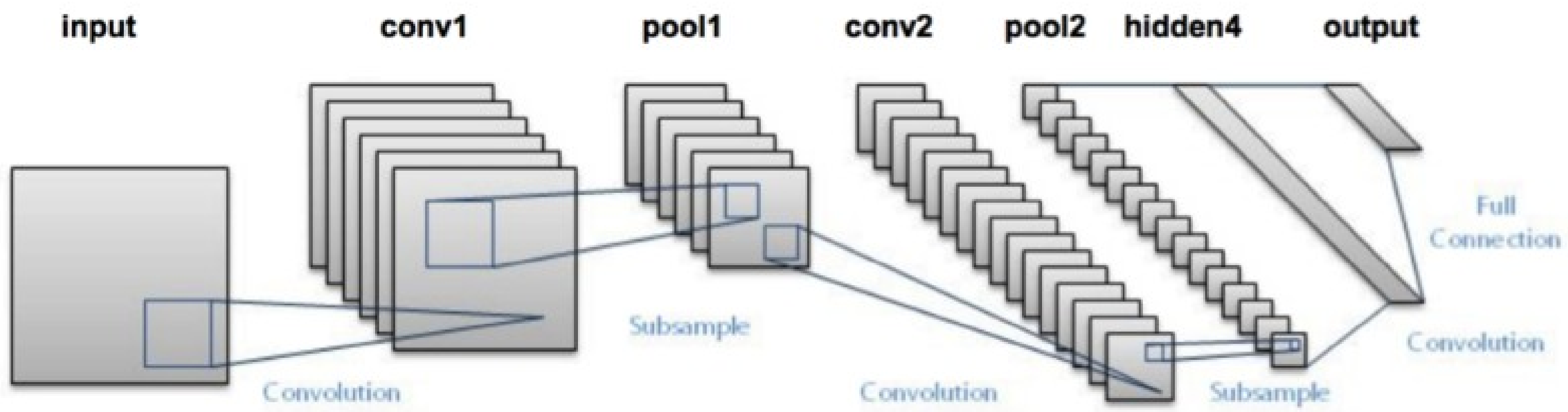


Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevs, Ilya Sutskever, Ruslan Salakhutdinov

http://jmlr.org/papers/volume15/srivastava14a.old/srivastava14a.pdf

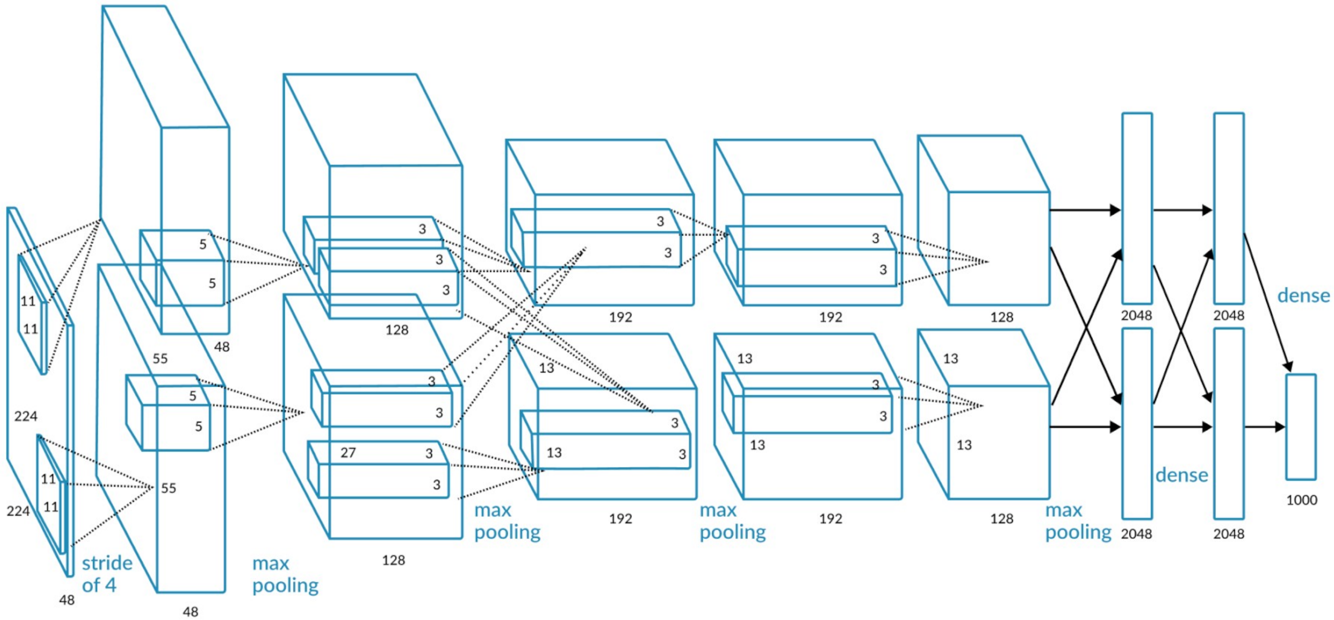
LeNet (LeCun 1998)



- premier réseau à convolutions
- appliqué à la reconnaissance de chiffres

Gradient-Based Learning Applied to Document Recognition
Yann LeCun, Léon Bottout, Patrick Haffner
http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

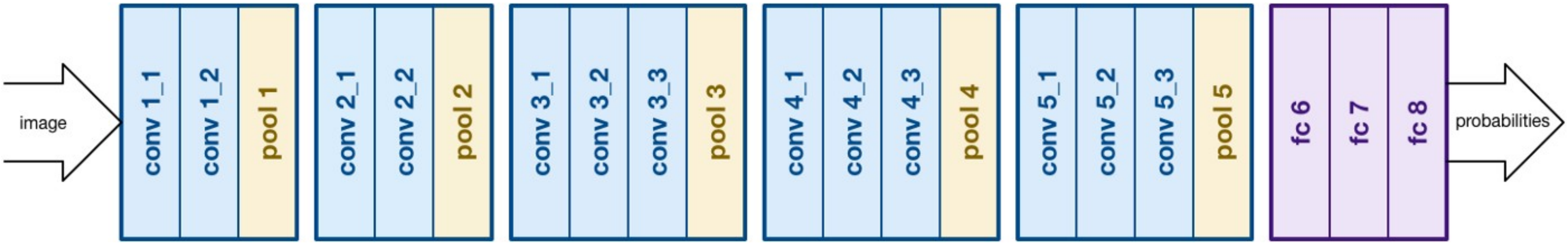
AlexNet (2012)



- présenté pour une compétition sur ImageNet : a réduit l’erreur de 26% à 15.3%
- premier CNN à obtenir des résultats meilleurs que les algos classiques sur ImageNet
- 2 lignes parallèles car entrainé sur 2 GPUs avec interconnections.

ImageNet Classification with Deep Convolutional Neural Networks
https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever

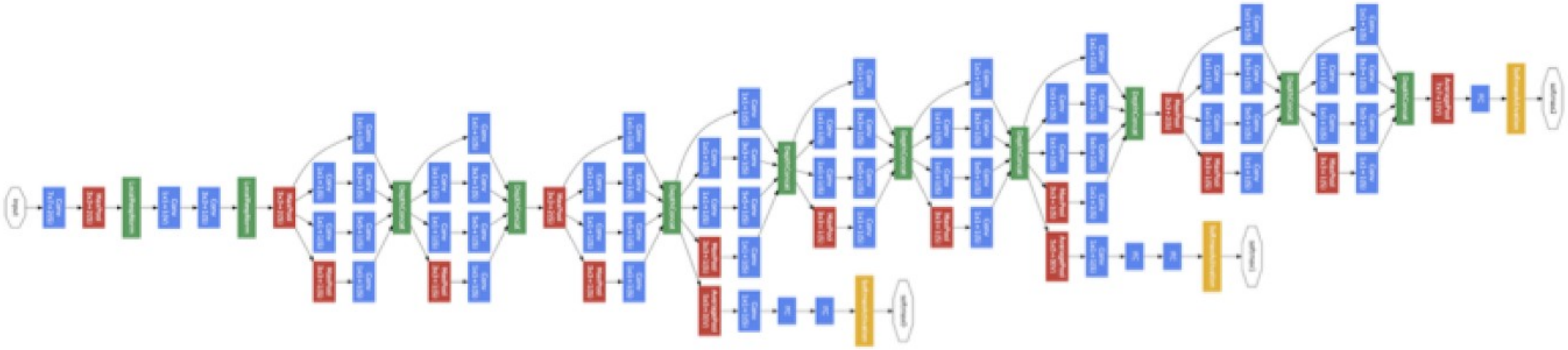
VGG (2014)



- développé par Simonyan and Zisserman (Visual Geometry Group à Oxford)
- nombre de filtres fortement augmenté

Very Deep Convolutional Networks For Large-Scale Image Recognition
Karen Simonyan, Andrew Zisserman
https://arxiv.org/pdf/1409.1556.pdf

GoogleNET - Inception (2014)



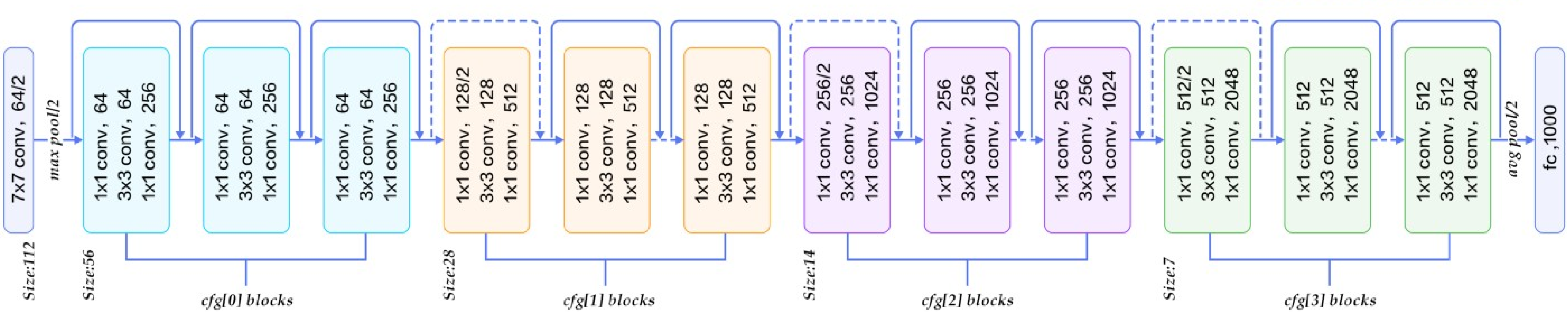
- module « Inception » répété
- basé sur des petites convolutions pour réduire le nombre de paramètres et donc augmenter la profondeur

Going deeper with convolutions

Christian Szegedy, Wei Liu, Chapel Hill, Yangqing Jia, Pierre Sermanet, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich

https://arxiv.org/pdf/1409.4842.pdf

ResNet (2015) réseau résiduel



- idée générale : créer des connexions directes pour « sauter » des couches
- permet de limiter les effets du *gradient vanishing*

Deep Residual Learning for Image Recognition
Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun
https://arxiv.org/pdf/1512.03385.pdf

Détection et segmentation

- **R-CNN** : Region of Interest (RoI)
  → **FastR-CNN**, **FasterR-CNN**
- **Mask-CNN** : segmentation
- **YOLO** : You Only Look Once