la valeur moyenne de la fenêtre). Un pooling très courant est un max pooling avec kernel de taille 2, stride de taille 2 et sans padding, comme présenté figure 1c.

1.3 Architectures convolutionnelles courantes

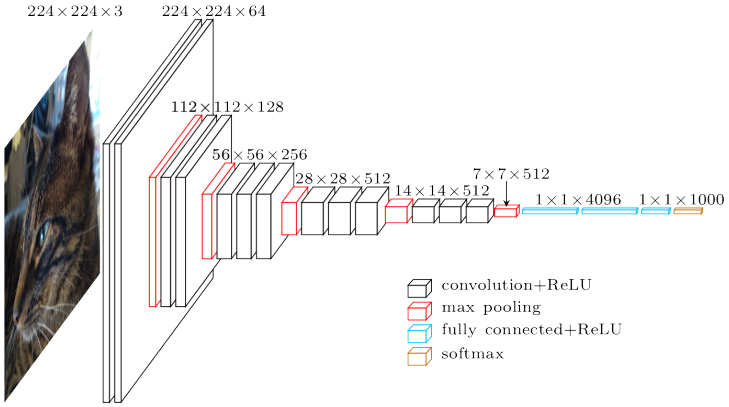

FIGURE 2 – Réseau VGG16.

Les réseaux de neurones convolutionnels classiques sont généralement composés d'une succession de couches de convolutions (avec ReLU) avec de plus en plus de filtres, et dont la dimension spatiale est progressivement réduite par des couches de max pooling possiblement jusqu'à aggregation totale des dimensions spatiales, il ne reste donc plus que la "profondeur" correspondant au nombre de filtres appliqués par la dernière convolution ($1 \times 1 \times C$). On y ajoute enfin généralement une ou quelques couches linéaires (appelées *fully-connected*). Un exemple de ce type d'architecture est le réseau VGG16 (Simonyan & Zisserman, 2014), *c.f.* Figure 2.

Depuis, des architectures plus complexes se sont développées, notamment les architectures Inception ou ResNet, et leurs dérivés et combinaisons.

1.4 Questions

1. Considérant un seul filtre de convolution de padding p, de stride s et de taille de kernel k, pour une entrée de taille $x \times y \times z$ quelle sera la taille de sortie ?
   Combien y a-t-il de poids à apprendre ?
   Combien de poids aurait-il fallu apprendre si une couche fully-connected devait produire une sortie de la même taille ?
2. ★ Quels avantages apporte la convolution par rapport à des couches fully-connected ? Quelle est sa limite principale ?
3. ★ Quel interêt voyez-vous à l'usage du pooling spatial ?
4. ★ Supposons qu'on essaye de calculer la sortie d'un réseau convolutionnel classique (par exemple celui en Figure 2) pour une image d'entrée plus grande que la taille initialement prévue (224×224 dans l'exemple). Peut-on (sans modifier l'image) calculer tout ou une partie des couches du réseau ?
5. Montrer que l'on peut voir les couches fully-connected comme des convolutions particulières.
6. Supposons que l'on remplace donc les fully-connected par leur équivalent en convolutions, répondre à nouveau à la question 4. Si on peut calculer la sortie, quelle est sa forme et son intérêt ?
7. On appelle champ récepteur (*receptive field*) d'un neurone l'ensemble des pixels de l'image dont la sortie de ce neurone dépend. Quelles sont les tailles des receptive fields des neurones de la première et de la deuxième couche de convolution ? Pouvez-vous imaginer ce qu'il se passe pour les couches plus profondes ? Comment l'interpréter ?