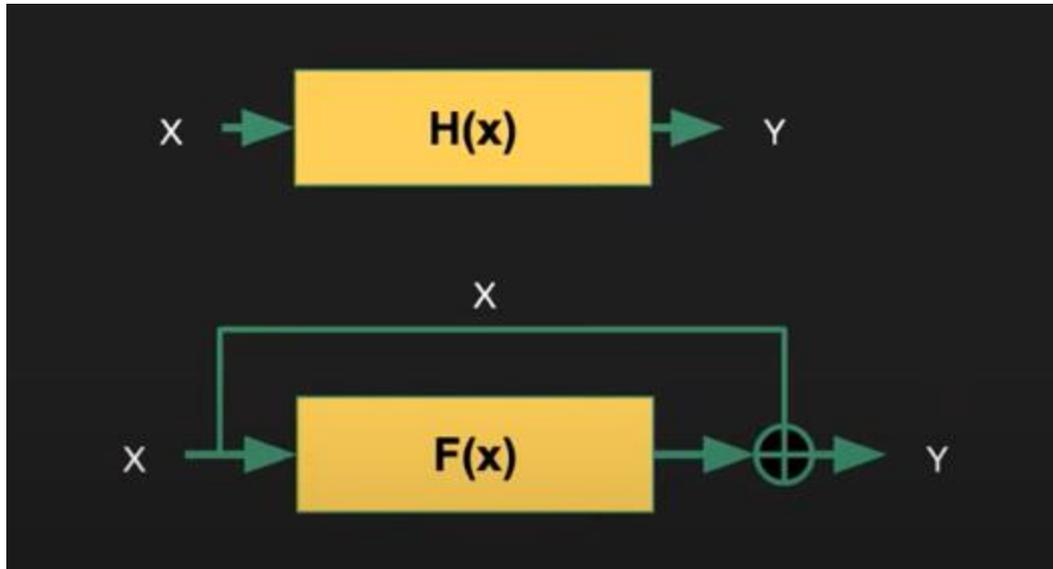


Les réseaux de neurones résiduels

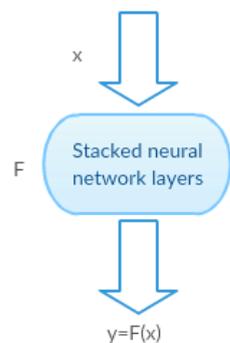
Un réseau neuronal résiduel appelé “ResNet” est un réseau neuronal artificiel réputé. Il s’assemble sur des constructions obtenues à partir des cellules pyramidales du cortex cérébral. Les réseaux neuronaux résiduels accomplissent cela en utilisant des raccourcis ou des “sauts de connexion” pour se déplacer sur différentes couches.



Initialement, le mappage souhaité est $H(x)$. Cependant, nous laissons les réseaux s’adapter à la cartographie résiduelle $F(x) = H(x) - x$, car le réseau a trouvé plus facile d’optimiser la cartographie résiduelle plutôt que la cartographie d’origine.

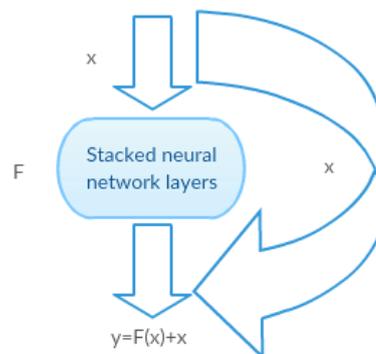
Les experts mettent en œuvre des modèles traditionnels de réseaux neuronaux résiduels avec des sauts à deux ou trois couches contenant des normalisations par lots et des non-linéarités entre les deux. Les scientifiques profitent également d’une matrice de poids supplémentaire pour apprendre les poids de saut dans certains cas. Le terme utilisé pour décrire ce phénomène est “**Highwaynets**”. Les modèles constitués de plusieurs sauts parallèles sont appelés “**Densenets**”. Les réseaux non résiduels peuvent également être appelés réseaux simples lorsqu’on parle de réseaux neuronaux résiduels.

Réseau de neurones simple



Hard to get $F(x)=x$ and make $y=x$
an identity mapping

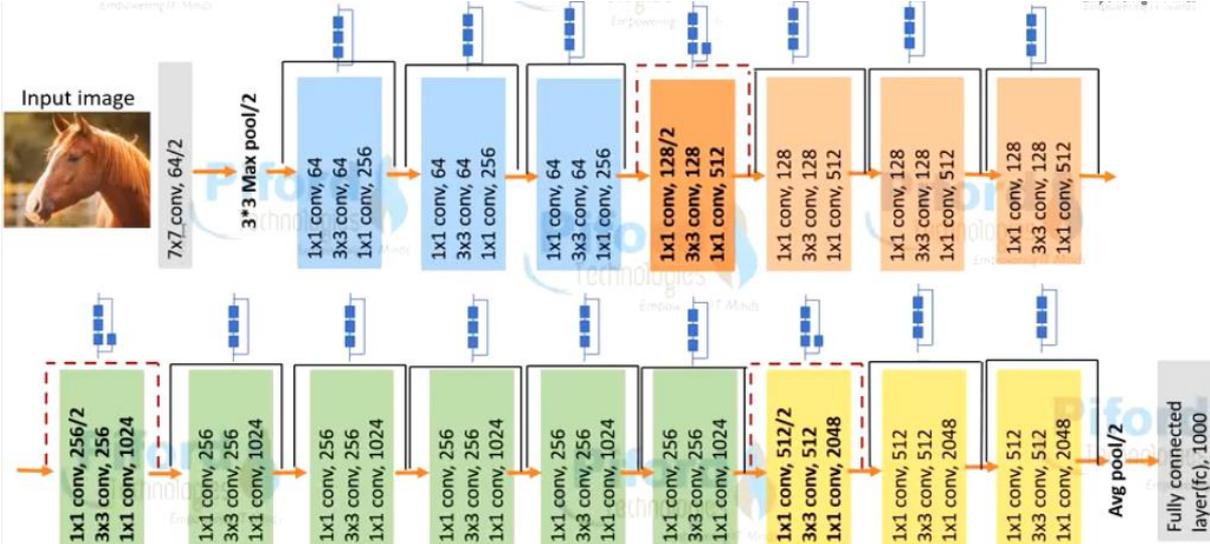
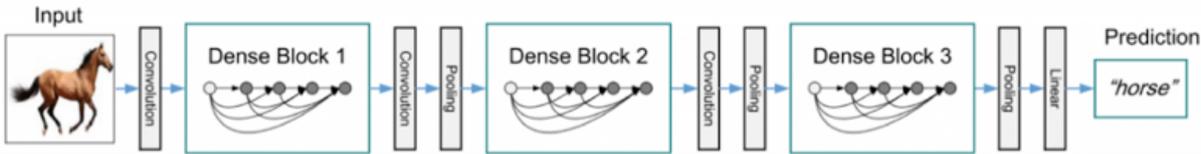
Réseau de neurones résiduel



Easy to get $F(x)=0$ and make $y=x$
an identity mapping

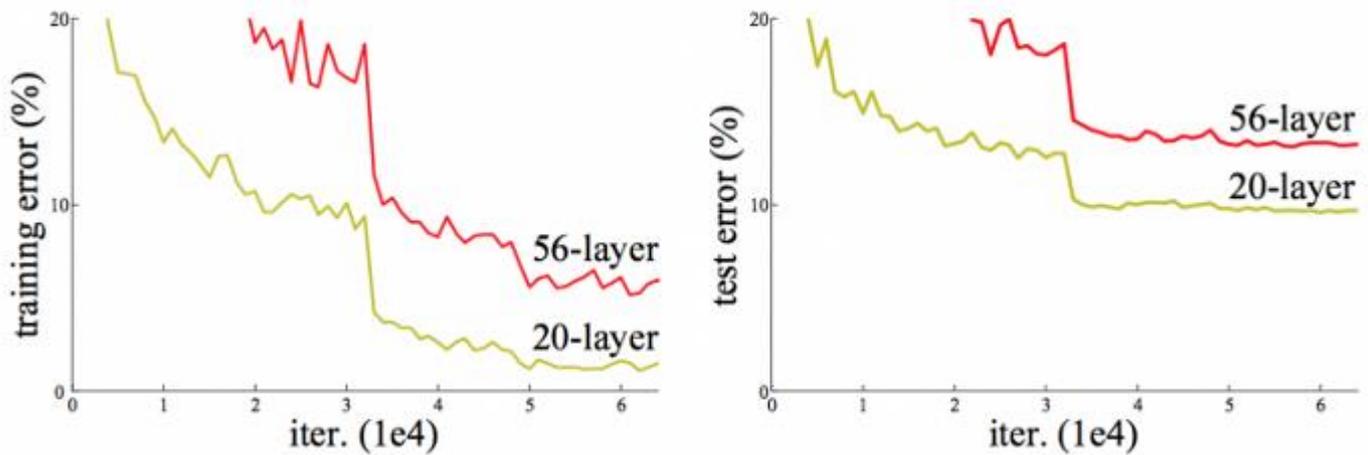
Mappage d’identité dans les blocs résiduels

De nombreuses applications de vision par ordinateur ont profité des fortes capacités de représentation du réseau neuronal résiduel et ont remarqué une augmentation massive. La classification d'images n'était pas la seule application de vision par ordinateur à utiliser ResNet – la reconnaissance des visages et la détection d'objets ont également bénéficié de cette innovation révolutionnaire.



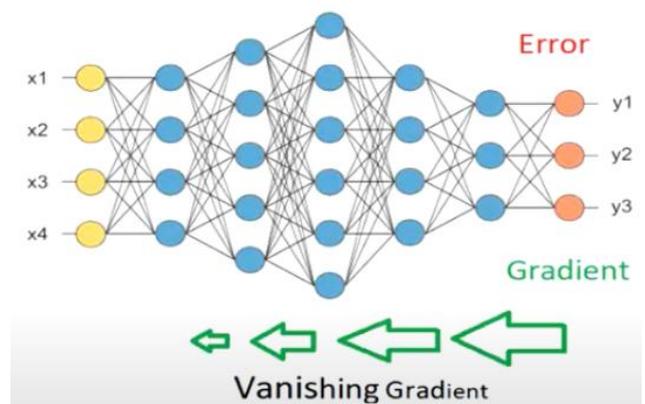
PROBLEME DE DEGRADATION

Lorsque des réseaux plus profonds commencent à converger, un problème de dégradation a été exposé: avec l'augmentation de la profondeur du réseau, la précision est saturée puis se dégrade rapidement.

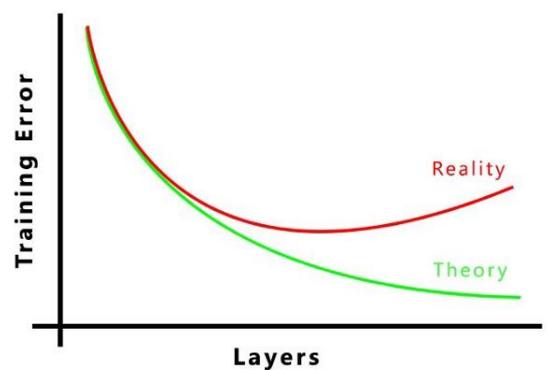


Il est observé que le réseau ayant un nombre de couches plus élevé (56 couches) entraîne une erreur d'entraînement plus élevée contrairement au réseau ayant un nombre de couches beaucoup plus faible (20 couches) entraînant ainsi des erreurs de test plus élevées

Avec l'augmentation de la profondeur du réseau, la précision sature (les réseaux apprennent tout avant d'atteindre la couche finale) puis commence à se dégrader rapidement si d'autres couches sont introduites.



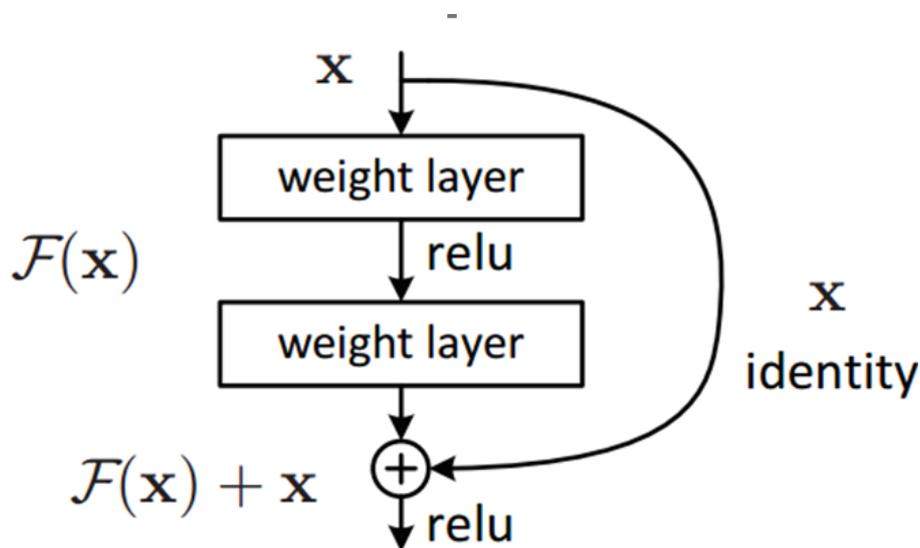
Théorie vs Réalité sur la variation de l'erreur d'entraînement avec une augmentation des couches,



Comment ResNet résout le problème du gradient de disparition (Vanishing Gradient)

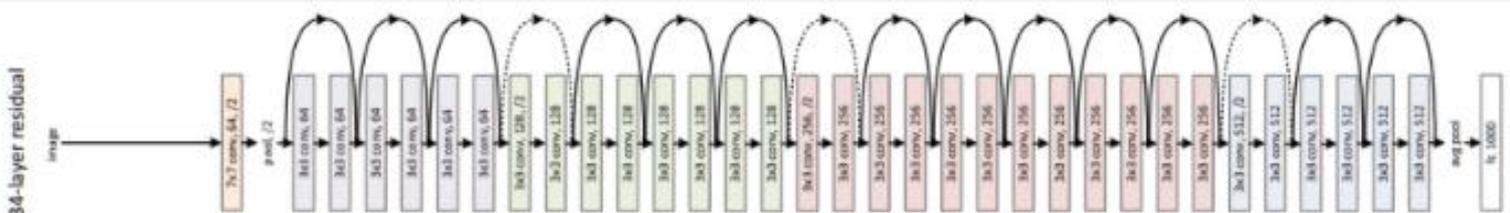
l'équipe Microsoft Research ont présenté un cadre d'apprentissage résiduel (ResNets) pour aider à faciliter la formation des réseaux qui sont beaucoup plus profonds qu'auparavant en éliminant le problème de dégradation. Ils ont prouvé avec des preuves que les ResNets sont plus faciles à optimiser et peuvent avoir une grande précision à des profondeurs considérables.

Le modèle saisit le concept de retenir les couches utiles et de ne pas utiliser celles qui ne sont pas utiles. Le modèle convertira ces dernières en cartes d'identité. C'est un facteur important du succès du réseau neuronal résiduel, car il est incroyablement simple de créer des couches correspondant à la fonction d'identité :



En outre, le fait qu'il soit possible de masquer les couches qui ne sont pas utiles est extrêmement utile. Une quantité massive de couches peut rendre les choses assez compliqués, mais avec l'aide des réseaux neuronaux résiduels, nous pouvons décider lesquelles nous voulons garder et lesquelles ne servent à rien.

L'intuition derrière la connexion de saut est qu'il est plus facile pour le réseau d'apprendre à convertir la valeur de $f(x)$ en zéro afin qu'il se comporte comme une fonction d'identité plutôt que d'apprendre à se comporter comme une fonction d'identité tout seul en essayant pour trouver le bon ensemble de valeurs qui vous donnerait le résultat.

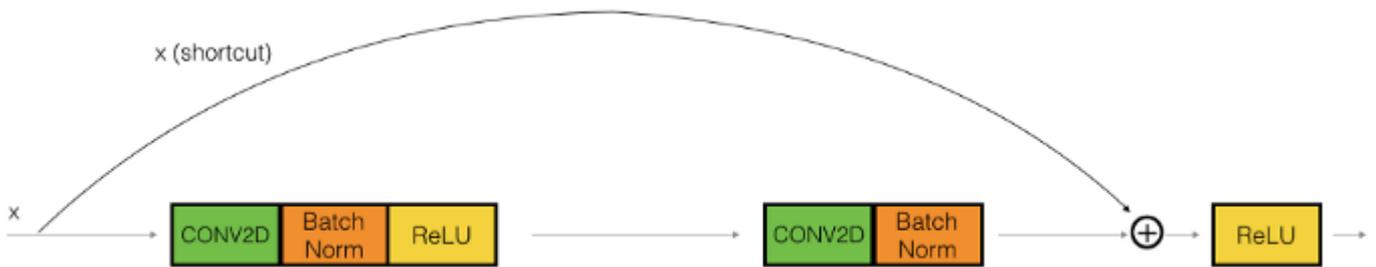


Modèle ResNet à 34 couches

ARCHITECTURE RESNET

ResNet utilise **deux blocs** de construction principaux pour construire l'ensemble du réseau.

1. Le bloc d'identité (comme ci-dessus) :

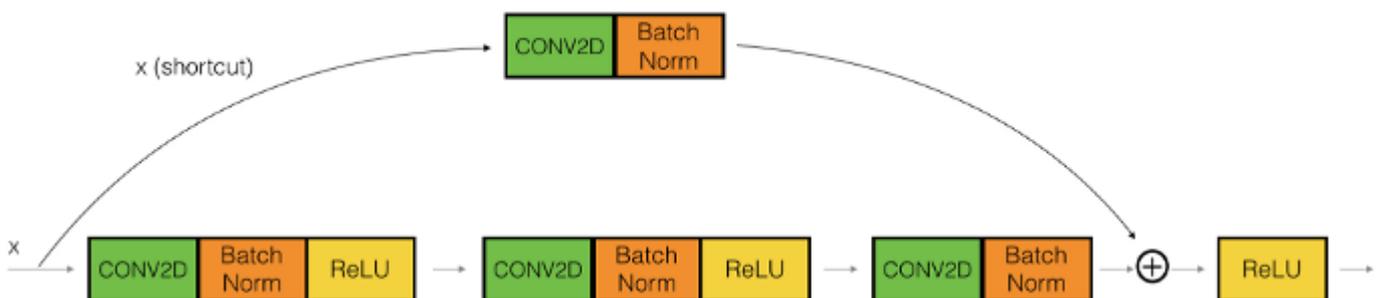


Les raccourcis d'identité (x) peuvent être directement utilisés lorsque l'entrée et la sortie sont de mêmes dimensions.

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}. \quad (1)$$

Fonction de bloc résiduel lorsque les dimensions d'entrée et de sortie sont identiques

2. Le bloc de convolution :



Lorsque les dimensions changent, A) Le raccourci effectue toujours le mappage d'identité, avec des entrées zéro supplémentaires complétées par la dimension augmentée. B) Le raccourci de projection est utilisé pour faire correspondre la dimension (faite par $1*1$ conv) en utilisant la formule suivante

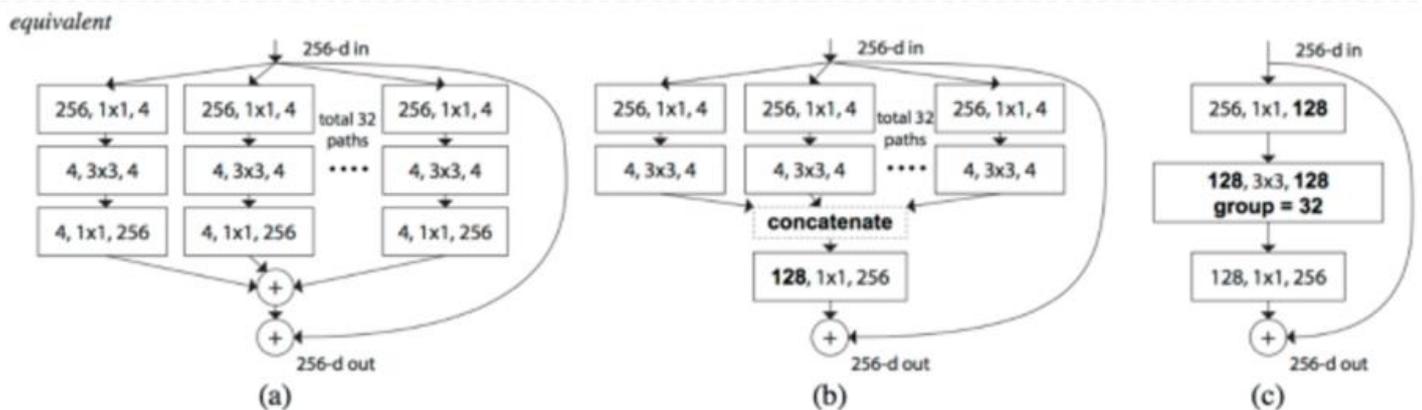
$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + W_s \mathbf{x}. \quad (2)$$

Fonction de bloc résiduel lorsque les dimensions d'entrée et de sortie ne sont pas les mêmes.

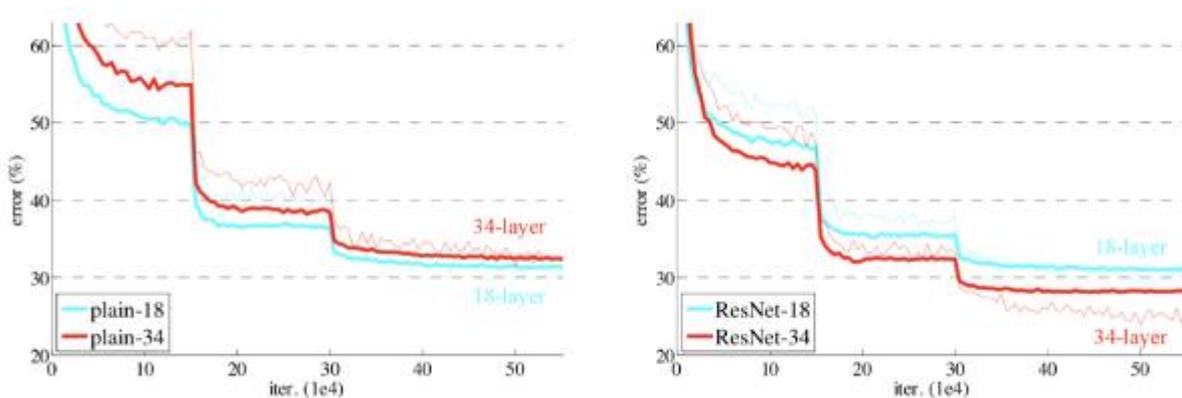
Le bloc conv permet de modifier et de restructurer les données entrantes afin que la sortie de la première couche corresponde aux dimensions de la troisième couche afin qu'elles puissent être ajoutées.

Ces composants permettent d'obtenir une optimisation et une précision supérieures pour les modèles d'apprentissage en profondeur.

L'image ci-dessous montre plusieurs interprétations d'un bloc résiduel :



Les résultats montrent avec précision l'effet de l'utilisation de ResNet sur des couches simples dans le graphique ci-dessous.

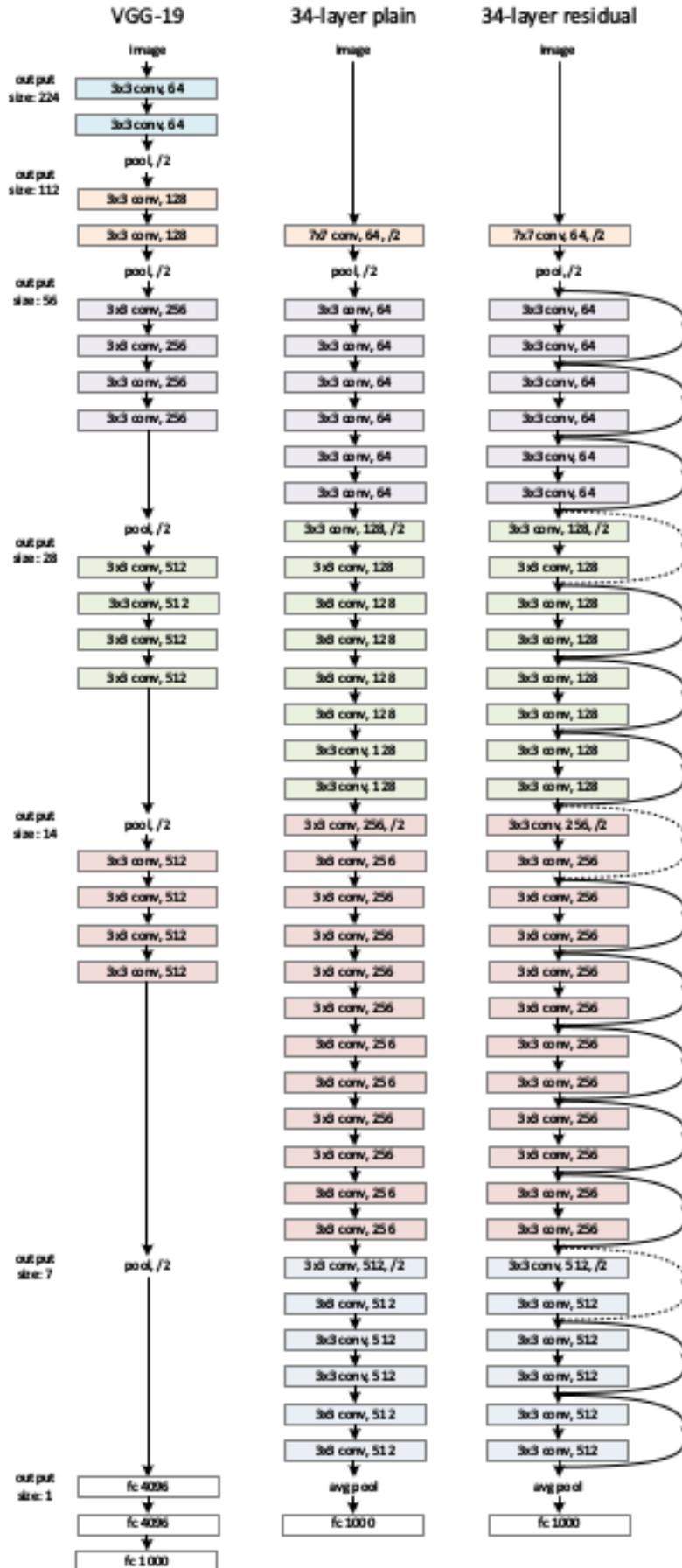


ResNet fonctionne mieux que les modèles de réseaux de neurones ordinaires, Même si le réseau à 18 couches n'est que le sous-espace d'un réseau à 34 couches, il fonctionne toujours mieux. ResNet surperforme d'une marge significative si le réseau est plus profond

	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	25.03

plaine vs ResNet

CONCEPTION DU RESEAU



Réseau VGG

Réseau VGG Plus profonde

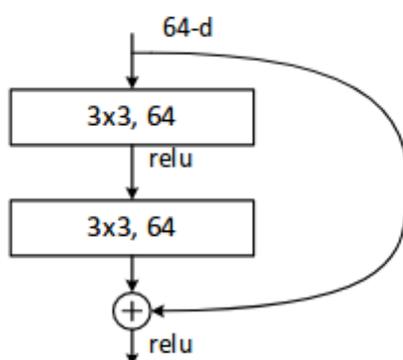
Réseau RESNET

Un réseau simple (réseau VGG type 18 couches) une variante plus profonde de celui-ci (34 couches) et des couches résiduelles (34 couches avec connexions résiduelles) qui utilise principalement des filtres 3 * 3. Échantillonnage vers le bas avec des couches CNN avec foulée 2. Couche de pooling moyenne mondiale et couche entièrement connectée à 1000 voies avec Softmax à la fin.

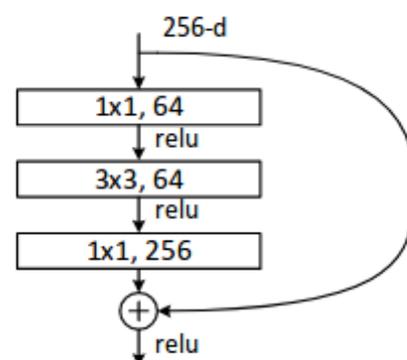
Les réseaux suivants sont étudiés :

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Chaque bloc ResNet a une profondeur de 2 couches (utilisé dans les petits réseaux comme ResNet 18, 34) ou une profondeur de 3 couches (ResNet 50, 101, 152) :



Bloc ResNet 2 couches



Bloc ResNet 3 couches

REMARQUES ET CONCLUSIONS

Le réseau ResNet converge plus rapidement que son homologue ordinaire.

Tous les blocs ResNet utilisent uniquement des raccourcis d'identité avec des raccourcis de projections utilisés uniquement lorsque les dimensions changent.

ResNet-34 a obtenu une erreur de validation dans le top 5 de 5,71% meilleure que BN-inception et VGG. ResNet-152 atteint une erreur de validation top-5 de 4,49%. Un ensemble de 6 modèles avec différentes profondeurs atteint une erreur de validation top-5 de 3,57%. **Gagner la 1ère place dans ILSVRC-2015**

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC' 14)	-	8.43 [†]
GoogLeNet [44] (ILSVRC' 14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	19.38	4.49

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except [†] reported on the test set).

ResNet en ImageNet résultats-2015

Par conséquent, nous pouvons facilement conclure que ResNet est sans aucun doute une étape importante dans l'apprentissage en profondeur. Avec ses connexions raccourcies/connexions ignorées, il a permis à la communauté de l'apprentissage en profondeur de s'aventurer dans des modèles de réseaux de neurones plus profonds, ce qui nous a donné à son tour des résultats nettement meilleurs.