

Etat de l'art sur la classification d'images à grande échelle. Etudier les architectures récentes offrant les meilleures performances sur la base ImageNet.

Soutenance de l'UE Communication

Julien Dejasmin

24 janvier 2019

Responsable pédagogique : M.SICRE Ronan

Sommaire

1 ImageNet

2 L'évolution des architectures des CNN au travers de la compétition de ImageNet.

3 Conclusion

4 Références

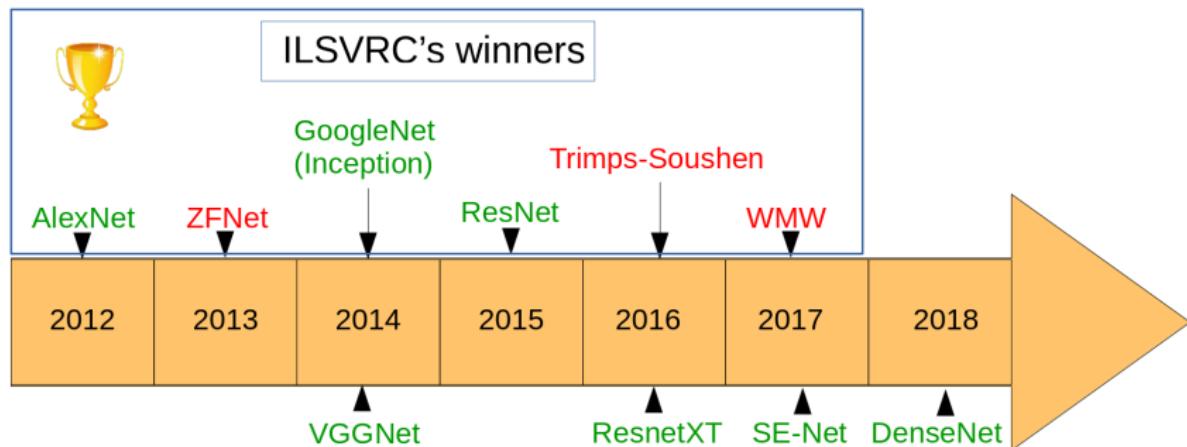
ImageNet

Histoire



- Collaboration de l'université de Stanford et de Princeton.
- Base de données : 15 millions d'images et 22.000 catégories.
- ImageNet Large Scale Recognition Vision Challenge (ILSVRC) :
 - Depuis 2010.
 - 1.000 images pour chacune des 1.000 catégories.
 - 50.000 images de validation.
 - 150.000 images de test.
 - Evaluation : top-1 et top-5.
- Objectif : faire avancer la recherche dans le domaine de la vision par ordinateur.

L'évolution des architectures des CNN au travers de la compétition de ImageNet.



Les gagnants de la compétition ILSVRC.

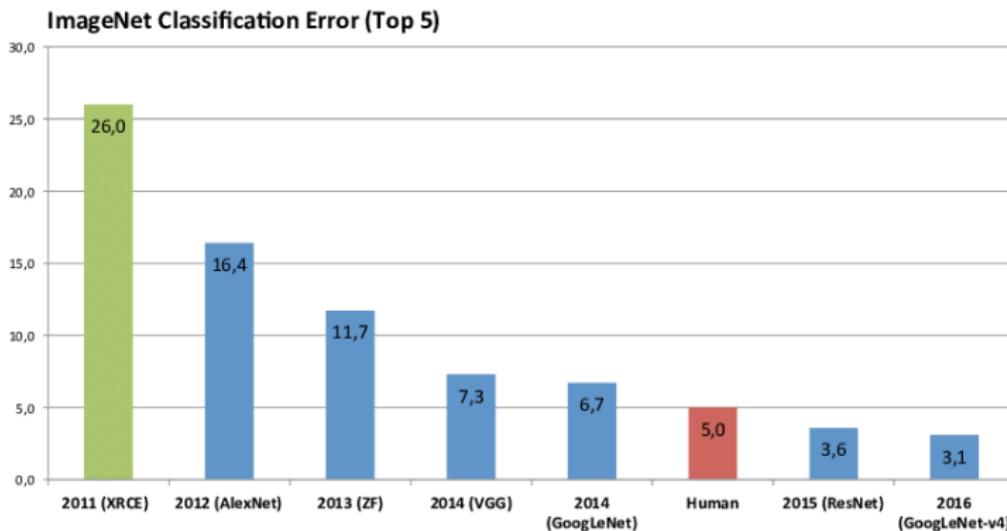


Figure: ILSVRC's Winners

AlexNet (2012)

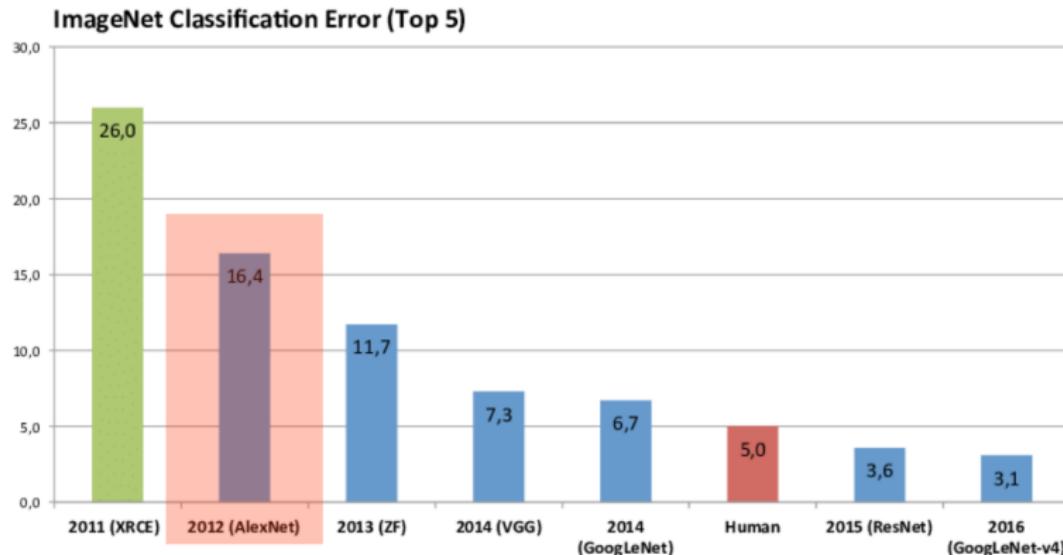


Figure: ILSVRC's Winners

AlexNet (2012)

- Auteur : Alex Krizhevsky
- 5 couches de convolution
- Points forts :
 - ReLu
 - Max pooling
 - Augmentation de données
 - Dropout
- Article associé: "ImageNet Classification with Deep Convolutional Neural Networks." [KSH12]

AlexNet (2012)

Architecture

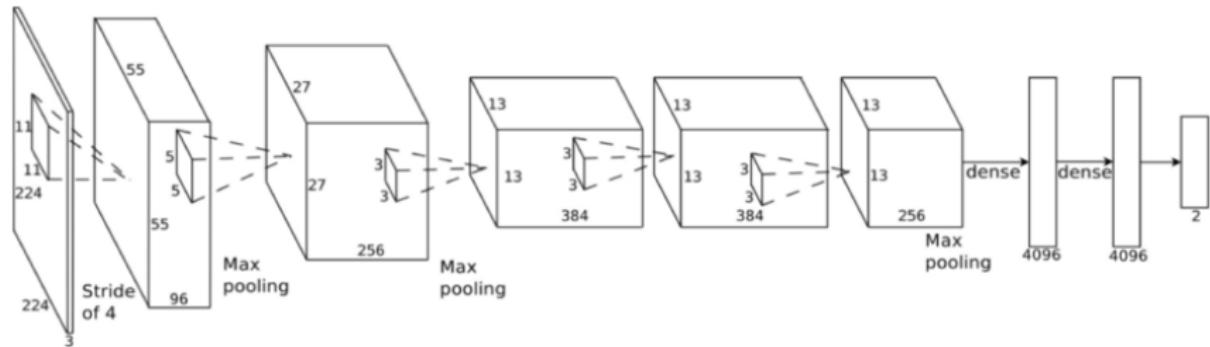


Figure: Architecture du CNN AlexNet.

ZFNet

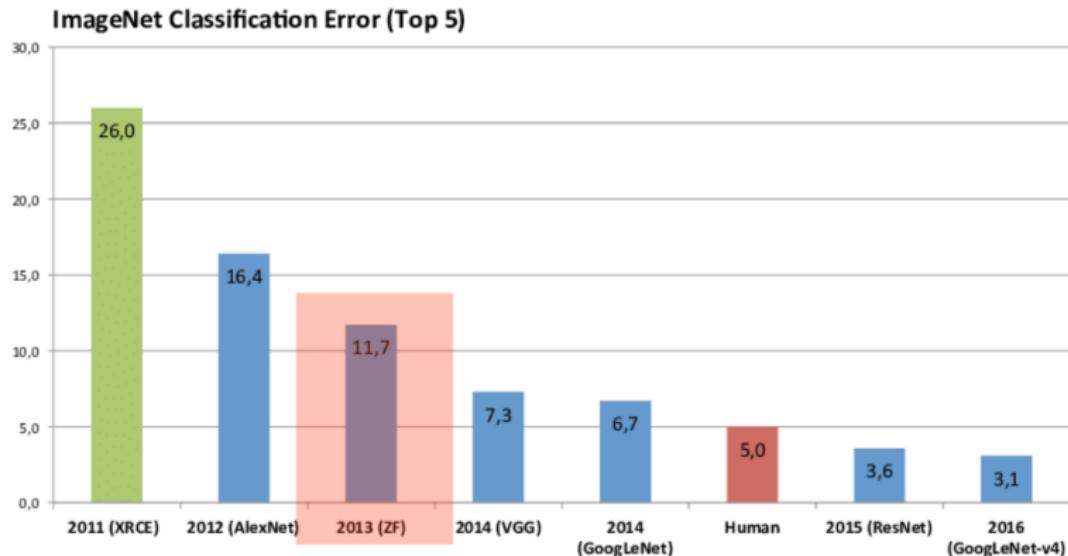


Figure: ILSVRC's Winners

VGGNet (2014)

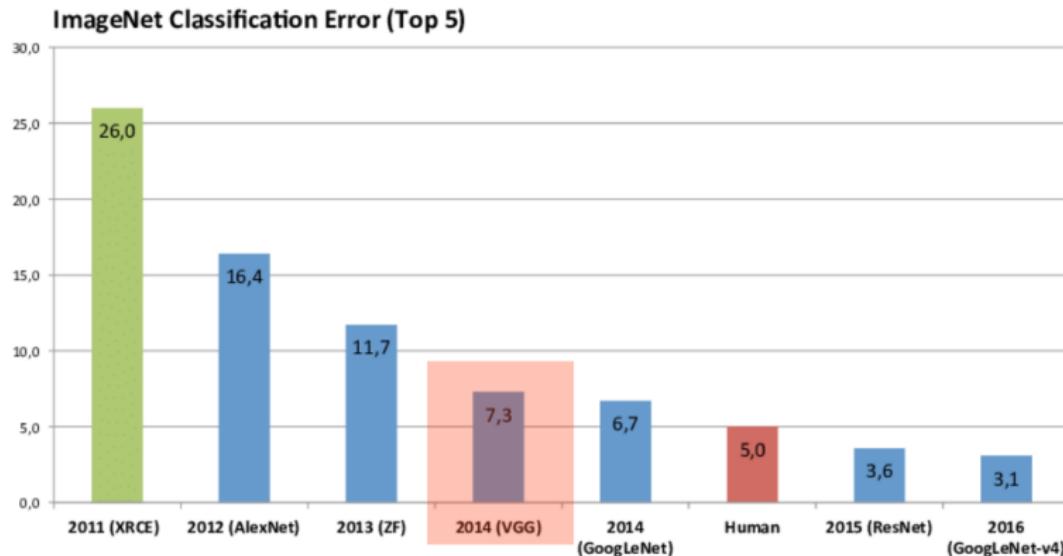


Figure: ILSVRC's Winners

VGGNet (2014)

- Auteurs : Visual Geometry Group de Oxford
- 2 versions :
 - VGG16 : 16 couches de convolutions
 - VGG 19 : 19 couches de convolutions
- Points forts :
 - Concept de bloc/modules
 - Remplace les noyaux de grandes taille en plusieurs noyaux de petites taille.
 - Diminution du nombre de paramètres.

VGGNet (2014)

Architecture

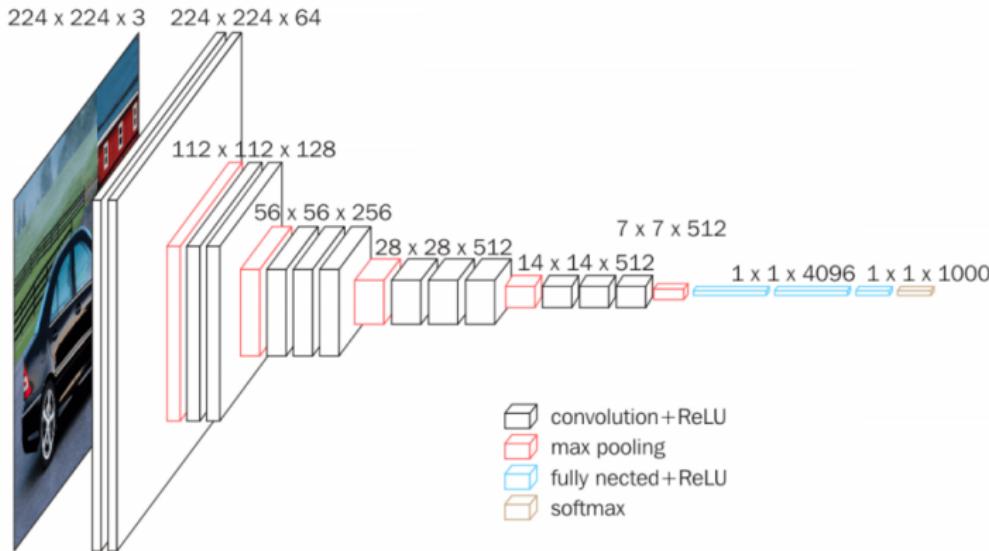


Figure: Architecture du CNN VGG16

GoogleNet (2015)

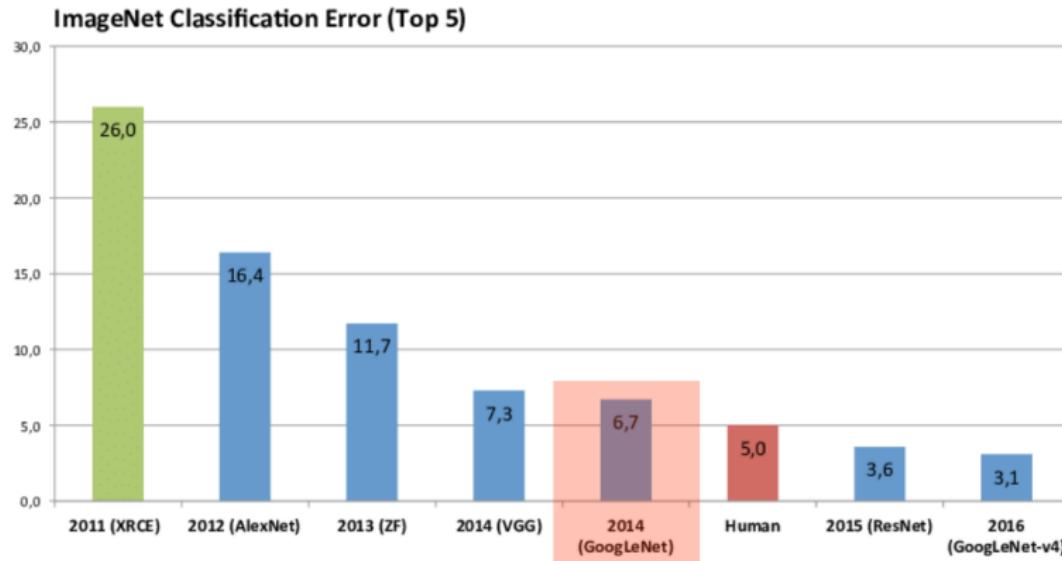


Figure: ILSVRC's Winners

GoogleNet (Inception) (2015)

- Auteur : Google.
- 22 couches.
- 12 fois moins de paramètres que AlexNet.
- Nouveau concept : module Inception.
- Article associé : "Going deeper with convolutions". [SLJ⁺15]

GoogleNet (Inception) (2015)

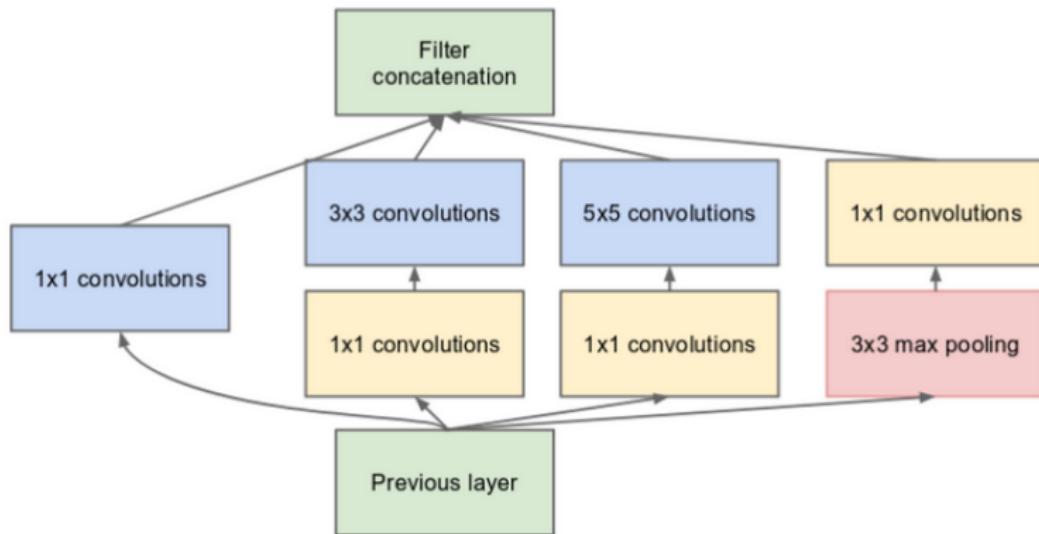


Figure: Module Inception

GoogleNet (Inception) (2015)

Architecture

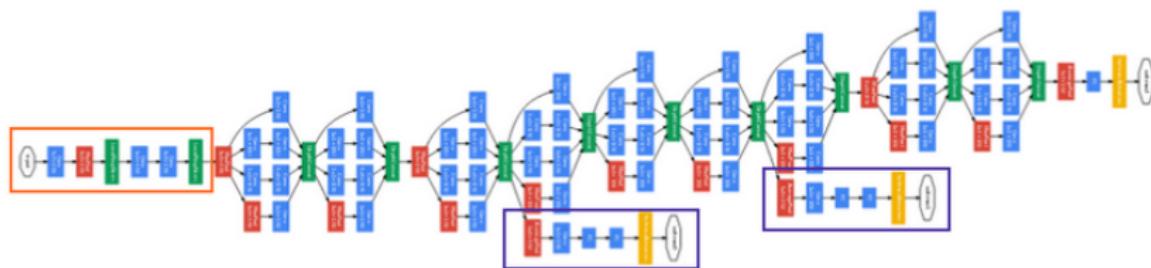


Figure: Architecture du CNN GoogleNet.

GoogleNet (Inception) (2015)

Les différentes version de Inception

⇒ La constante évolution de Inception à conduit à différentes versions.

- Inception v1.
- Inception v2.
- Inception v3.
- Inception v4.
- Inception-ResNet v1.
- Inception-ResNet v2.
- Article : "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning". [SIV16]

ResNet (2015)

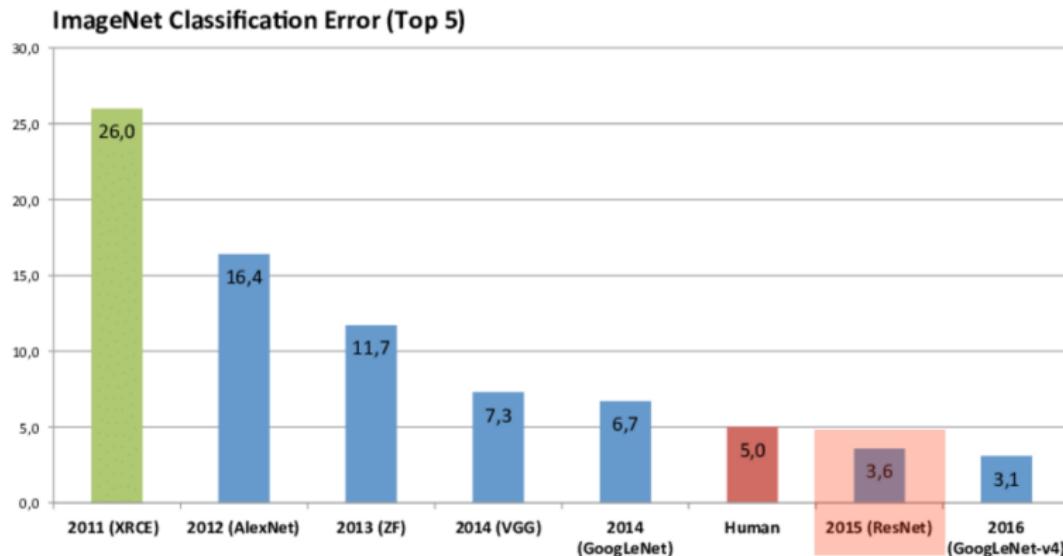


Figure: ILSVRC's Winners

ResNet (2015)

- Auteur : Microsoft Research Asia.
- 152 couches !
- idée principale : introduire une connexion de raccourci qui saute une ou plusieurs couches.
- Article associé : "Deep Residual Learning for Image Recognition".[HZRS16]

ResNet (2015)

Bloc résiduel

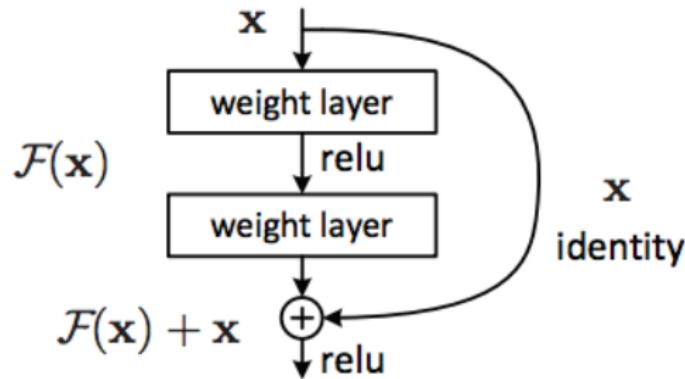


Figure: Bloc résiduel du CNN ResNet.

ResNet (2015)

Architecture

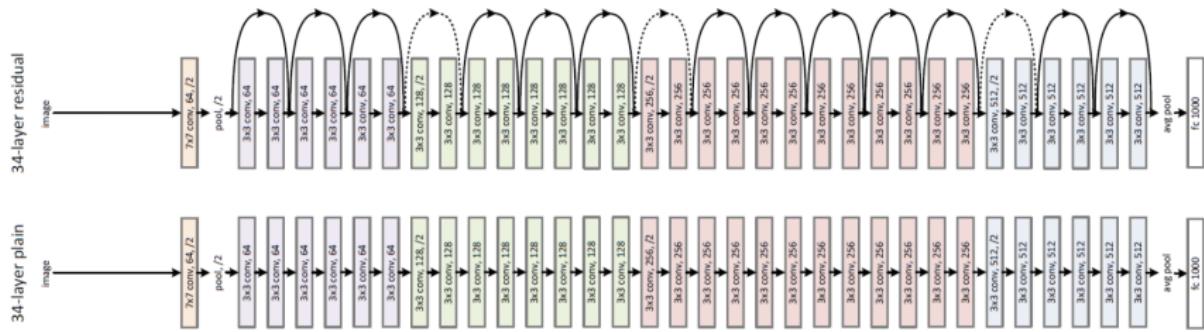


Figure: Architecture du CNN ResNet.

Wide Residual Networks (2016)

- Élégant et simple.
- Article associé : "Wide Residual Networks". [ZK16]

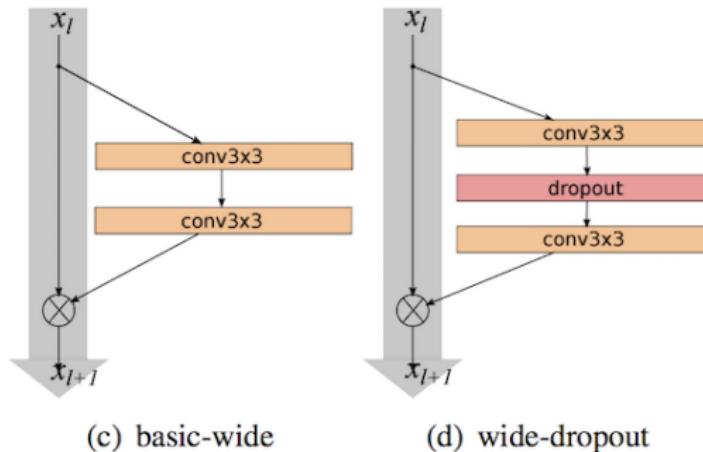
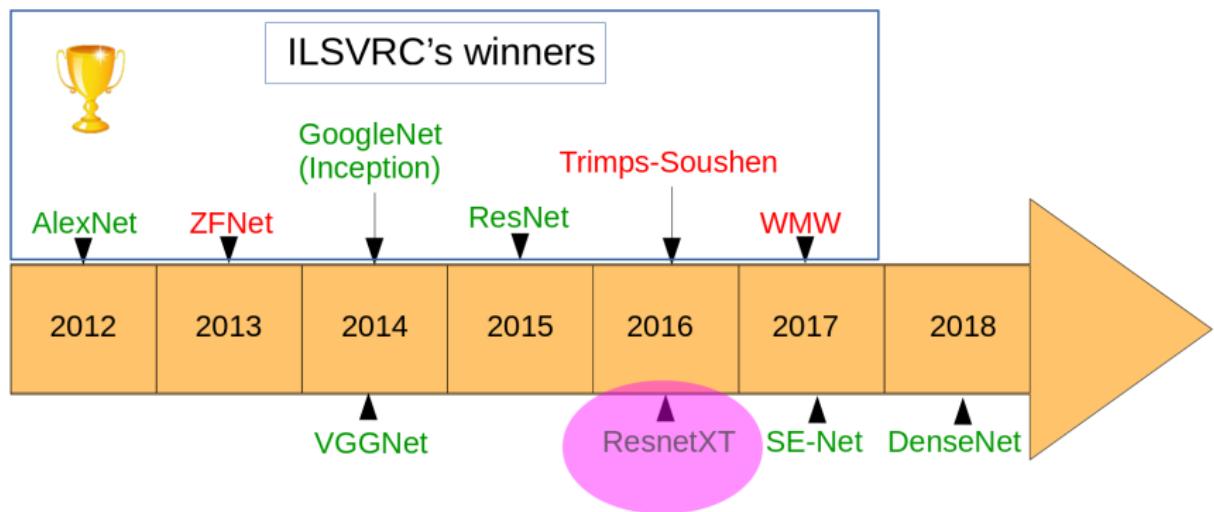


Figure: Différents blocs résiduels utilisé dans [ZK16]

ResNetXT (2016)



ResNetXT (2016)

Architecture

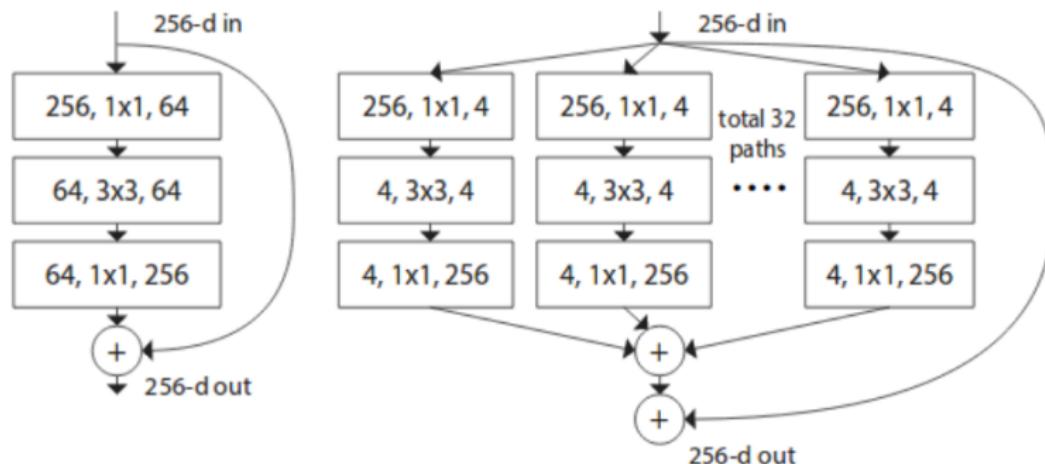


Figure: Comparaison des modules des CNN ResNet (à gauche) et ResNetXT (à droite).

ResNetXT (2016)

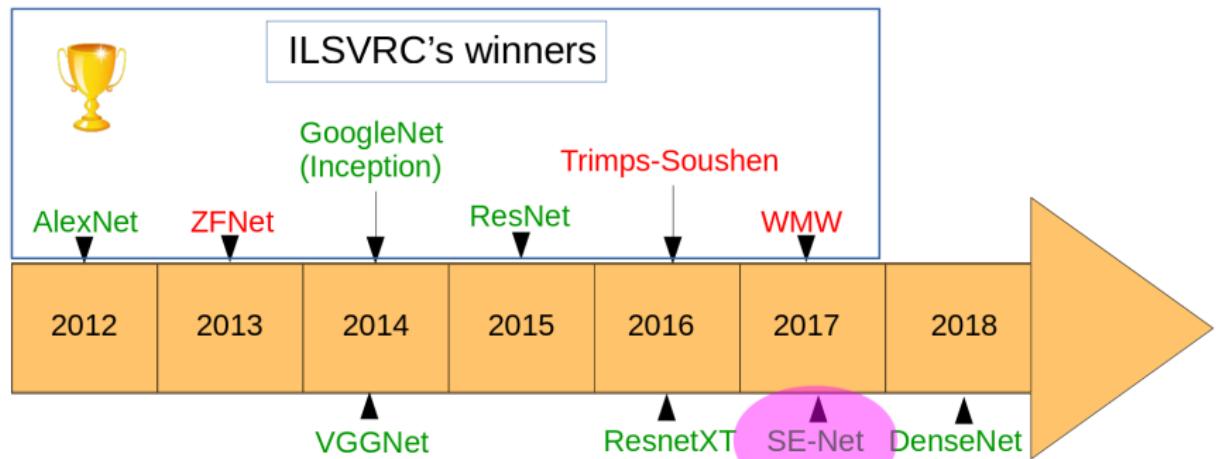
- Améliore les performances :

	setting	top-1 err (%)	top-5 err (%)
<i>1 × complexity references:</i>			
ResNet-101	1 × 64d	22.0	6.0
ResNeXt-101	32 × 4d	21.2	5.6
<i>2 × complexity models follow:</i>			
ResNet- 200 [15]	1 × 64d	21.7	5.8
ResNet-101, wider	1 × 100d	21.3	5.7
ResNeXt-101	2 × 64d	20.7	5.5
ResNeXt-101	64 × 4d	20.4	5.3

Figure: Résultats des architectures Resnet et ResnetXT sur le dataset ImageNet.

- Article associé : "Aggregated Residual Transformations for Deep Neural Networks".[XGD⁺17]

SE-Net (2018)



SE-Net (2018)

- Idée principale : ajouter un bloc aux CNN déjà existant :
 - Les CNN utilisent des filtres de convolution pour extraire des informations hiérarchique des images.
 - Fusion des informations spatiales et des informations des différents filtres (canaux).
 - Jusqu'ici les CNN pondèrent chaque canal de manière égale.
 - SE-NET veut changer cela.
- Article associé : "Squeeze and excitation network". [HSS18]

SE-Net (2018)

Module SE

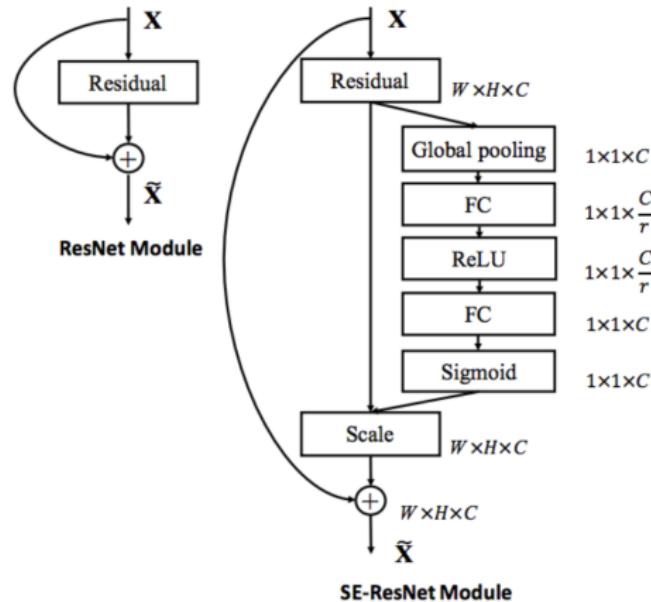


Figure: Module SE.

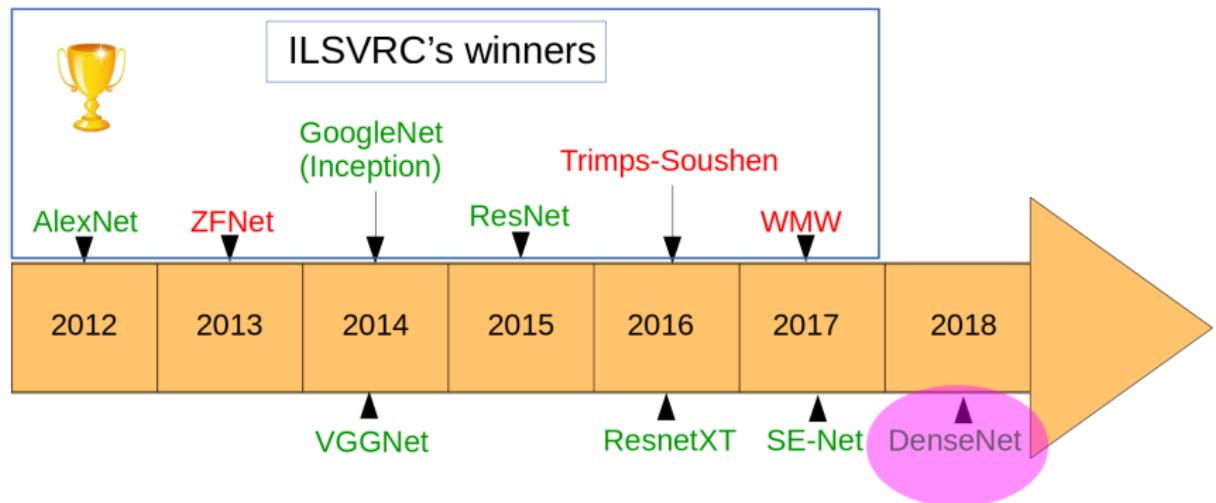
SE-Net (2018)

Résultats

	original		re-implementation			SENet		
	top-1 err.	top-5 err.	top-1 err.	top-5 err.	GFLOPs	top-1 err.	top-5 err.	GFLOPs
ResNet-50 [9]	24.7	7.8	24.80	7.48	3.86	23.29 _(1.51)	6.62 _(0.86)	3.87
ResNet-101 [9]	23.6	7.1	23.17	6.52	7.58	22.38 _(0.79)	6.07 _(0.45)	7.60
ResNet-152 [9]	23.0	6.7	22.42	6.34	11.30	21.57 _(0.85)	5.73 _(0.61)	11.32
ResNeXt-50 [43]	22.2	-	22.11	5.90	4.24	21.10 _(1.01)	5.49 _(0.41)	4.25
ResNeXt-101 [43]	21.2	5.6	21.18	5.57	7.99	20.70 _(0.48)	5.01 _(0.56)	8.00
BN-Inception [14]	25.2	7.82	25.38	7.89	2.03	24.23 _(1.15)	7.14 _(0.75)	2.04
Inception-ResNet-v2 [38]	19.9 [†]	4.9 [†]	20.37	5.21	11.75	19.80 _(0.57)	4.79 _(0.42)	11.76

Figure: Tableau comparatif des résultats obtenus avec les modules SE.

DenseNet (2018)



DenseNet (2018)

- Exploite l'idée des raccourcies de ResNet.
- Toutes les couches sont directement reliées entre elles.
- Assure un maximum de flux d'informations et de gradient.
- Article associé : "Densely Connected Convolutional Networks".[HLW16]

DenseNet (2018)

Architecture

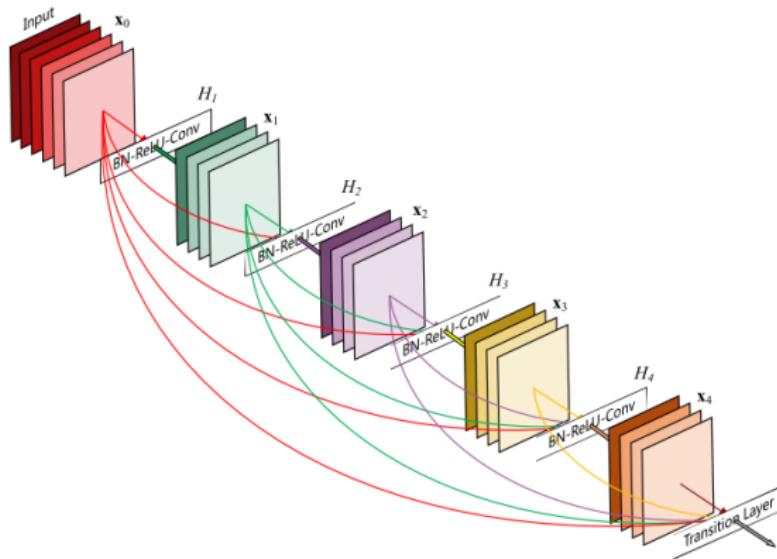


Figure: Architecture de DenseNet.

Conclusion

- Tour d'horizon des architectures de CNN qui ont marqué le domaine de la vision d'ordinateur à travers la classification d'un grand jeu de données réputé difficile : ImageNet.
- Si intéressé, possibilité d'obtenir la liste des papiers de cette présentation et d'en discuter autour d'un café (offert).

Références

-  Gao Huang, Zhuang Liu, and Kilian Q. Weinberger, *Densely connected convolutional networks*, arXiv preprint arXiv:1608.06993 (2016).
-  J. Hu, L. Shen, and G. Sun, *Squeeze-and-excitation networks*, IEEE Conference on Computer Vision and Pattern Recognition, 2018.
-  Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, *Deep residual learning for image recognition*, 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pp. 770–778.
-  Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, *Imagenet classification with deep convolutional neural networks*, 2012.
-  Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke, *Inception-v4, inception-resnet and the impact of residual connections on learning*, CoRR [abs/1602.07261](https://arxiv.org/abs/1602.07261) (2016).
-  Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott