



*Bienvenue!*

**ÉCOLE D'ÉTÉ FRANCOPHONE  
EN APPRENTISSAGE PROFOND**

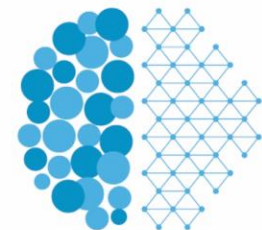
---

**21-25 août 2017**



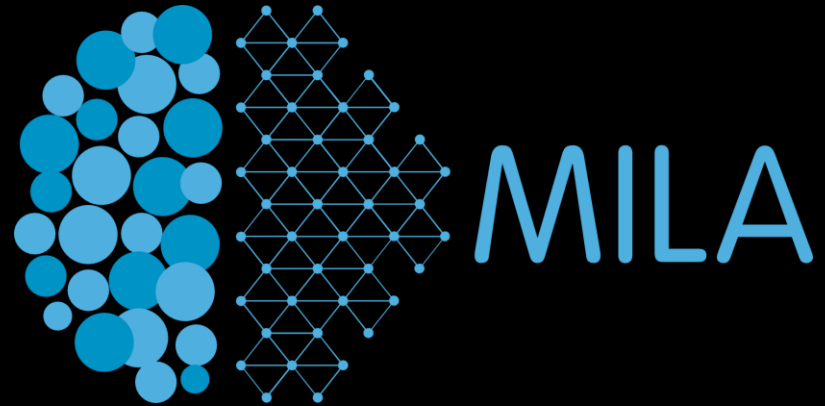
**IVADO**

HEC Montréal  
Polytechnique Montréal  
Université de Montréal



**MILA**

Institut  
des algorithmes  
d'apprentissage  
de Montréal



Alexei Nordell-Markovits  
[alexei@elementai.com](mailto:alexei@elementai.com)

Pedro Oliveira Pinheiro  
[pedro@elementai.com](mailto:pedro@elementai.com)

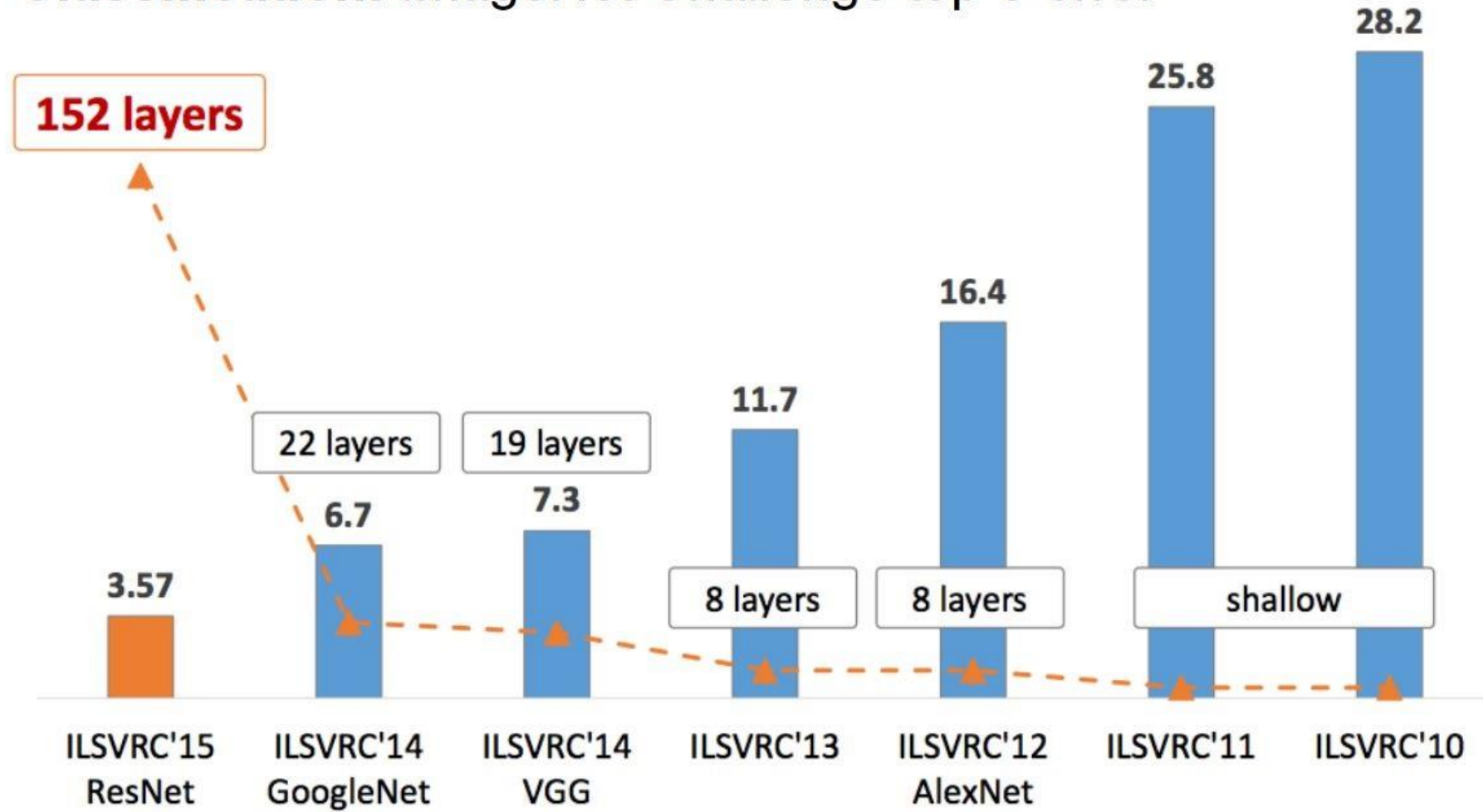
**ELEMENT<sup>AI</sup>**

# INTRODUCTION AUX RÉSEAUX CONVOLUTIONNELS (SUITE)

# Objectifs de la présentation

- ImageNet et réseaux modernes
- Divers

# Classification: ImageNet Challenge top-5 error



Source : [Kaiming He](#)

# ImageNet

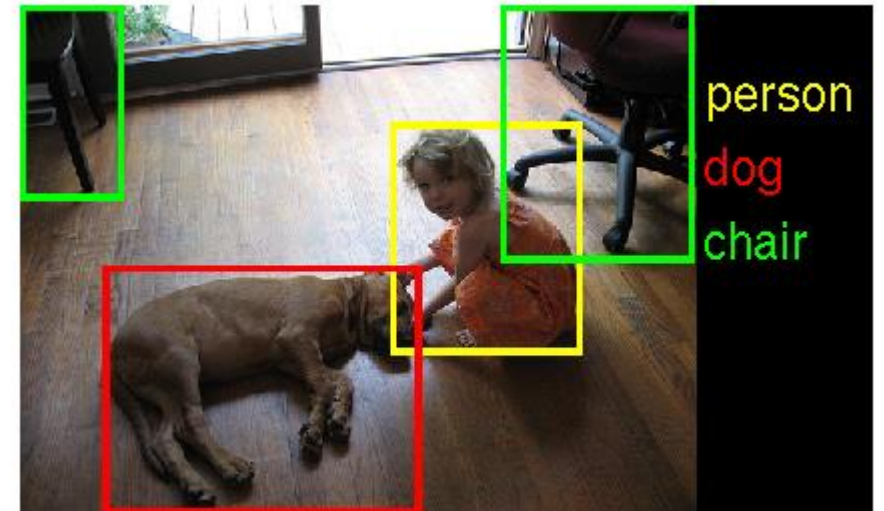
- Plus de 10 millions d'images
  - 3 millions d'images pour la V1 ( 2009 )
  - Organisées via la hiérarchie Wordnet
  - 2 ans (!?)
- Google + Mechanical Turk
  - Google pour trouver les images
  - Mechanical Turk pour valider la classification



Source : [http://www.image-net.org/papers/imagenet\\_cvpr09.pdf](http://www.image-net.org/papers/imagenet_cvpr09.pdf)

# ImageNet - ILSVRC

- Compétition annuelle
  - Métriques multiples
- *Classification task*
  - 1000 catégories
  - Top-5 pour chaque image
  - Pourcentage de succès



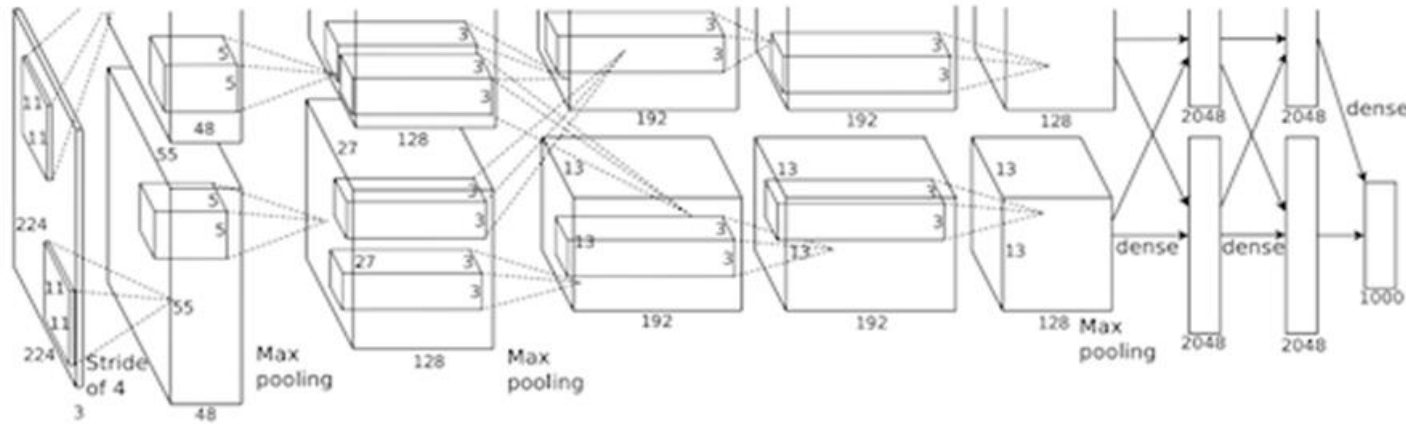
Source : <http://www.image-net.org/challenges/LSVRC/2015/>

# ImageNet





# 2012 - AlexNet - 16.4%

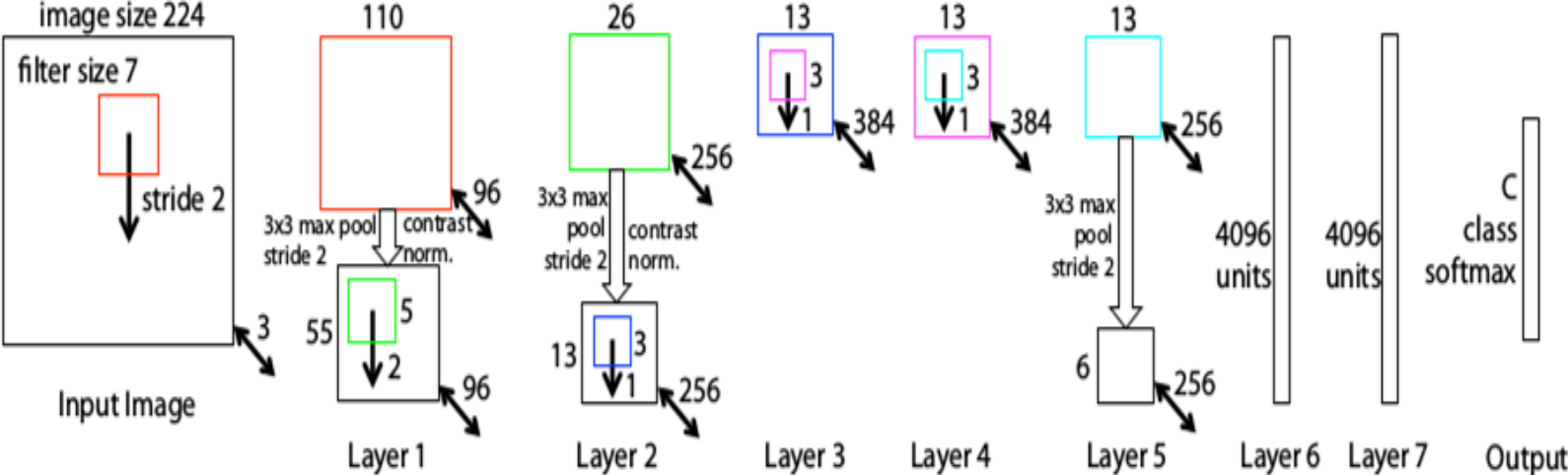


AlexNet architecture (May look weird because there are two different “streams”. This is because the training process was so computationally expensive that they had to split the training onto 2 GPUs)

- ReLU
- Dropout
- 2 GPU / 1 semaine

Source : [Adit Deshpande](#)

# 2013 - Clarifai/ZG - 11.7%



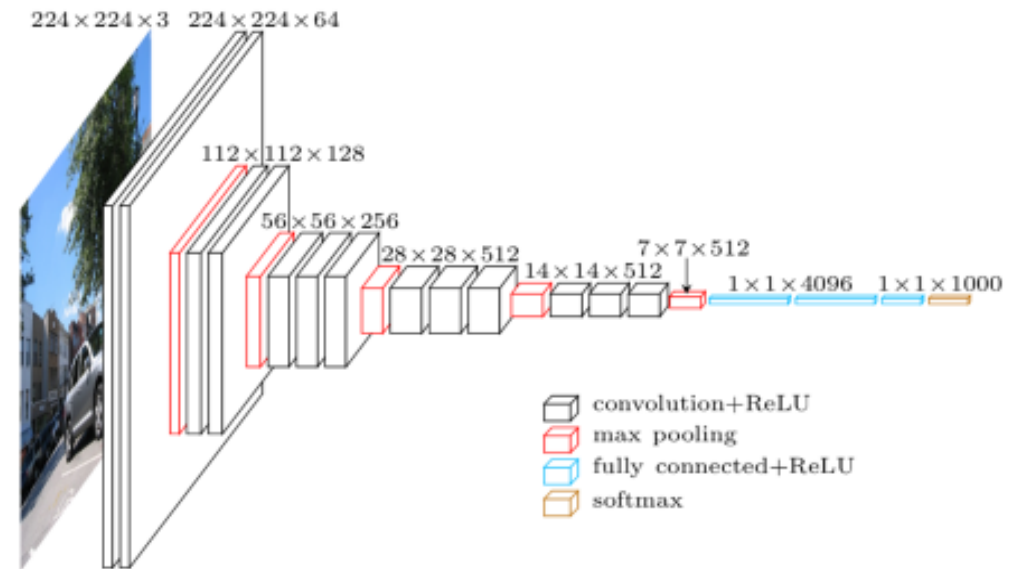
ZF Net Architecture

- Similaire à AlexNet
- “Plumbing”
- Deconvnet

Source : <https://arxiv.org/abs/1311.2901v3>

# 2014 - VGG (2<sup>e</sup>)

- Architecture simple
  - Composée de convolutions 3x3 et 2x2
  - 19 couches
  - 4 GPU – 2 semaines
- La profondeur est importante!
  - Pas besoin d'être complexe pour bien fonctionner
- Disponible dans [Caffe zoo](#)

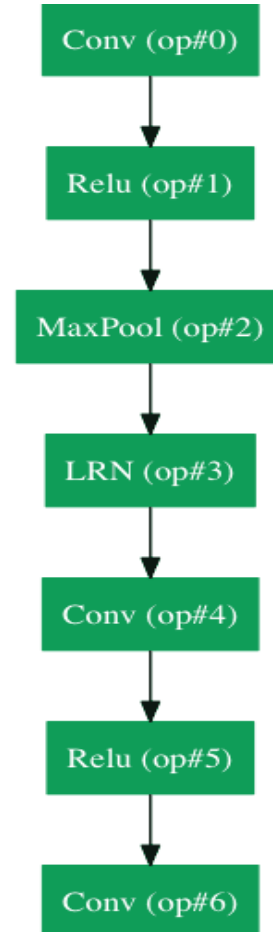


Source : <https://arxiv.org/abs/1409.1556>

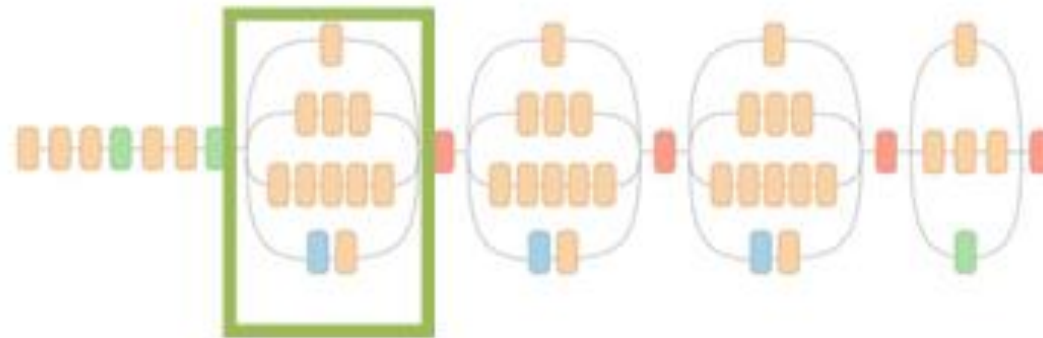
# 2014 - Inception - 7.3%

- Les réseaux convolutionnels ne sont pas si effrayants
  - Sauf Inception
- Mais... petit nombre de paramètres
  - 2-3 GPU – 1 semaines
- V1...V7

# 2014 - Inception - 7.3%



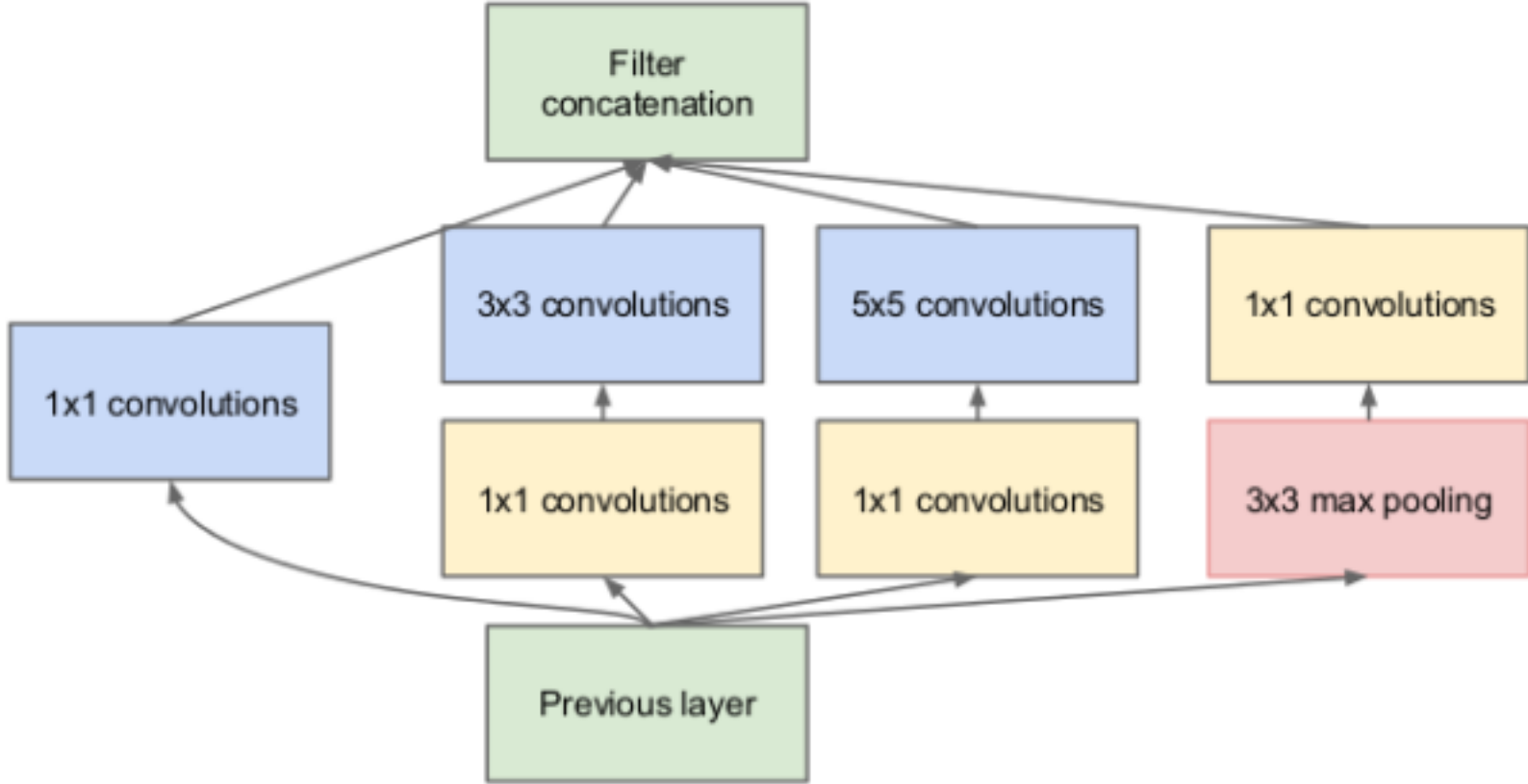
# 2014 - Inception



- Convolution
- AvgPool
- MaxPool
- Concat
- Dropout
- Fully connected
- Softmax

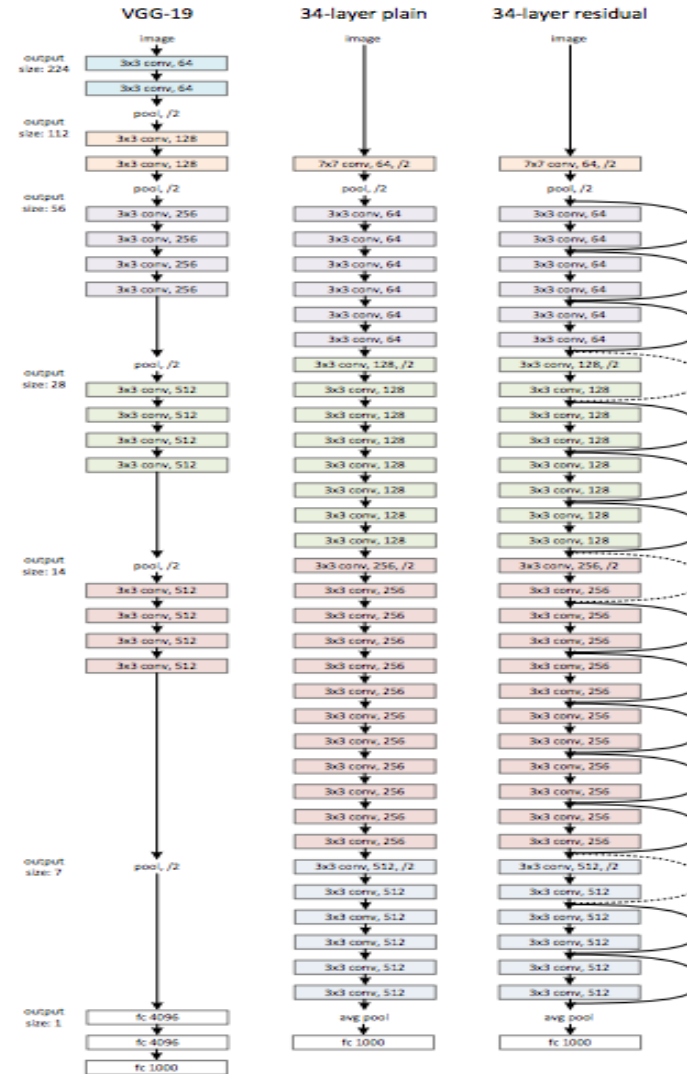
Green box shows parallel region of GoogLeNet

# 2014 - Inception



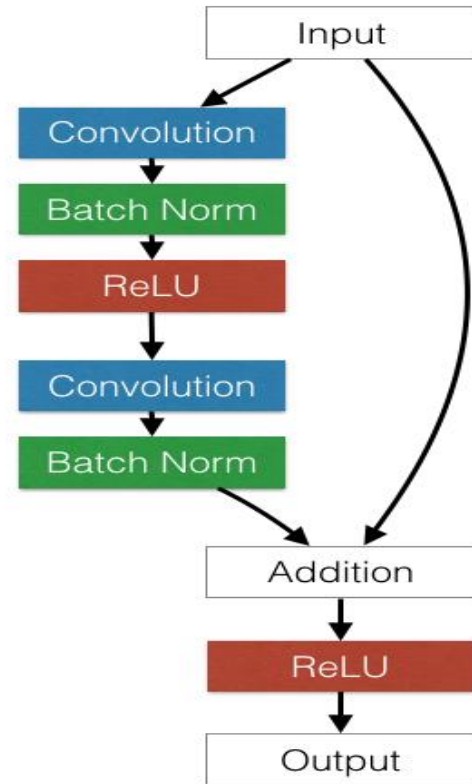
# 2015 - Resnet - 3.57%

- Meilleur qu'un humain!
- Quelques bémols
- 152 couches
- 8 GPU – 3 semaines





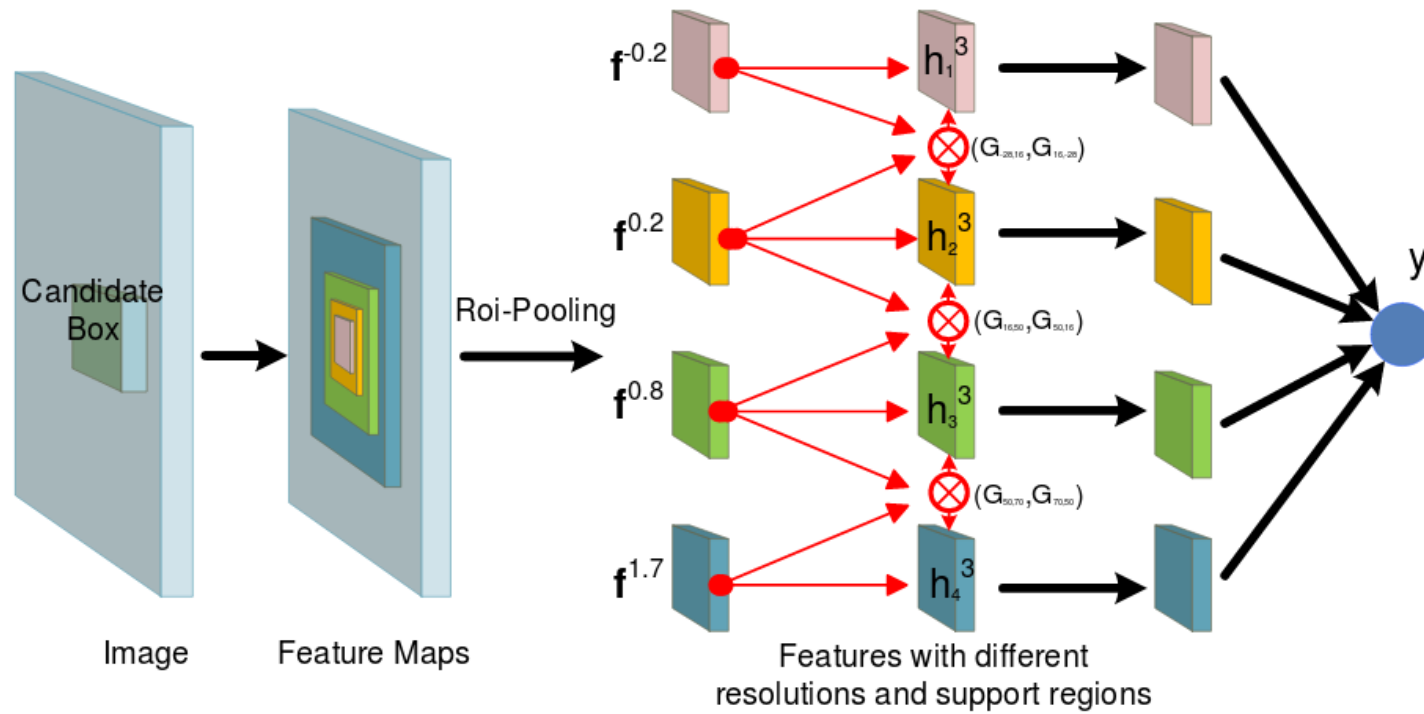
# 2015 - Resnet - module résiduel



- <https://github.com/facebook/fb.resnet.torch>

Source: <http://torch.ch/blog/2016/02/04/resnets.html>

# 2016 - Culmage



Source : <https://arxiv.org/pdf/1610.02579.pdf>

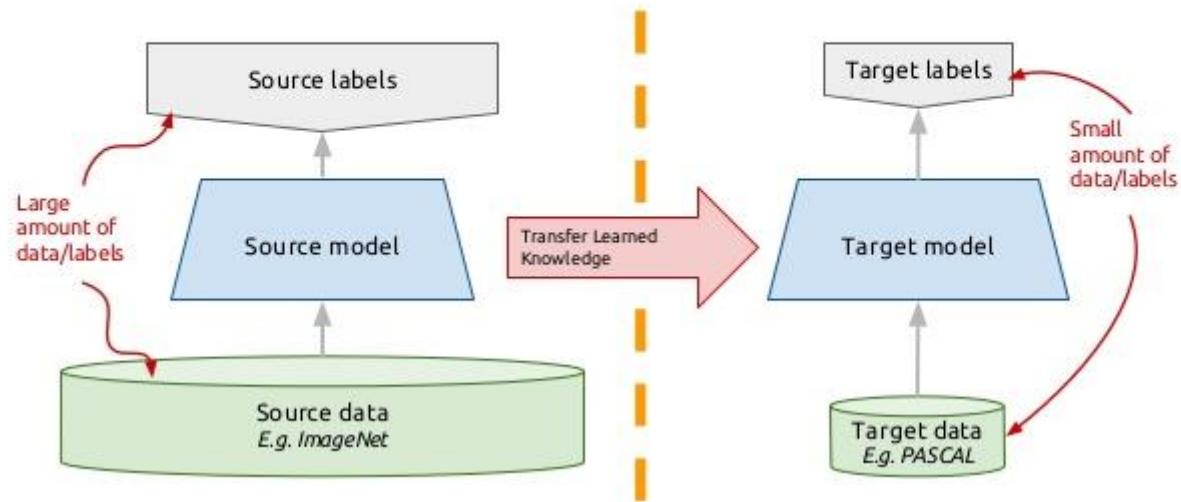
**2017 - C'est la fin!**

Vous pensez faire mieux qu'un CNN?

<http://cs.stanford.edu/people/karpathy/ilsvrc/>

# Apprentissage par transfert

## Transfer learning: idea



Un chat c'est un chat

# Apprentissage par transfert

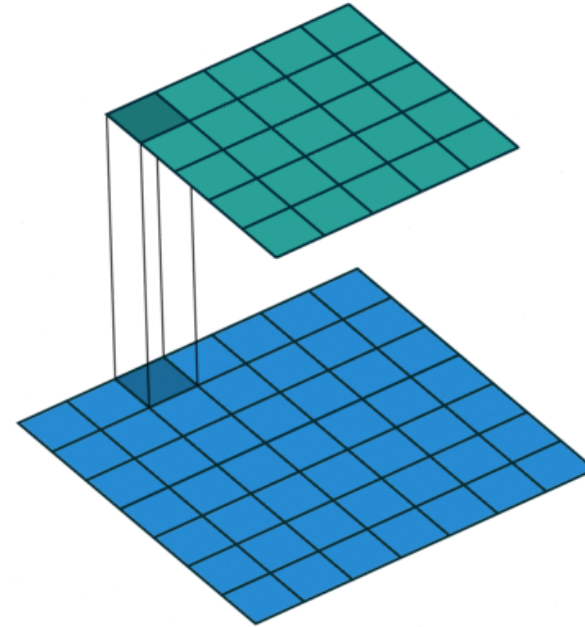
- Entraîner un CNN est long
  - Jours, parfois semaines
- Beaucoup de caractéristiques sont partagées
  - Pourquoi ne pas couper la couche de «décisions» et utiliser la nôtre
- Deux scénarios:
  - Utiliser un réseau convolutionnel comme outil pour définir les caractéristiques
  - Re-entraîner le réseau en partant des poids initiaux
- Voir : <https://github.com/pkmital/CADL/tree/master/session-4/libs>

# Apprentissage par transfert

- Inception v3
  - Utilisé dans quelques projets à l'interne
  - Comme *feature extractor* avec des arbres de décisions
  - Avec une phase d'entraînement supplémentaire connectée à une couche complète
  - Voir : [https://www.tensorflow.org/tutorials/image\\_retraining](https://www.tensorflow.org/tutorials/image_retraining)
  - Aussi : <https://research.googleblog.com/2016/03/train-your-own-image-classifier-with.html>

# Divers - 1 x 1 convolution

- L'intuition est difficile
  - Réduire la dimensionnalité
- Poids uniques
  - Donc transformation linéaire
  - Suivi de RELU



*In Convolutional Nets, there is no such thing as “fully-connected layers”. There are only convolution layers with 1x1 convolution kernels and a full connection table. – [Yann LeCun](#)*



# Divers – augmentation des données

- Plus de données = succès
  - Bruits
  - Coupes
  - Rotations
  - Translations
  - Etc

## Data Augmentation:

a. No augmentation (= 1 image)



b. Flip augmentation (= 2 images)



c. Crop+Flip augmentation (= 10 images)



73

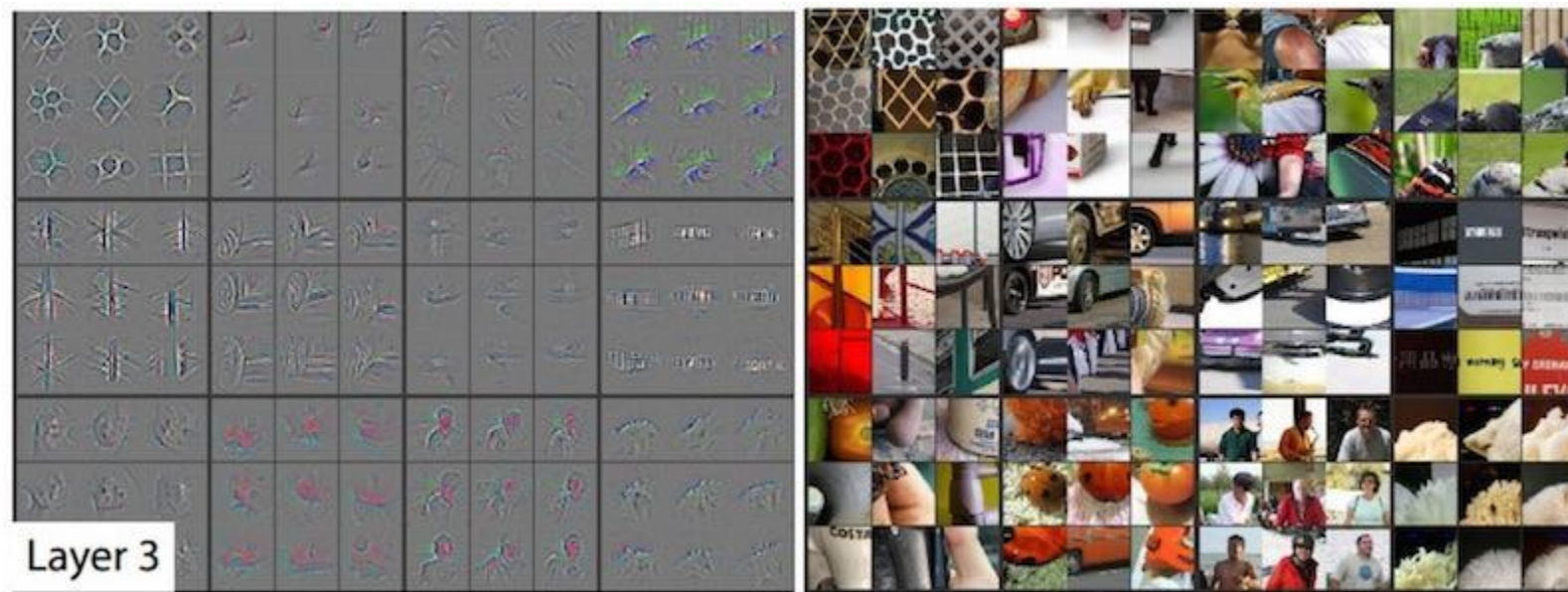
Source : <https://www.quora.com/Is-a-convolution-neural-network-immune-to-image-resolutions-If-so-what-makes-this-possible>

# Divers

- *Dropout*
  - Fonctionne également pour les réseaux convolutionnels
  - [Tensorflow without a Phd](#)
- *Ensemble methods*
  - Pourquoi voter une fois?

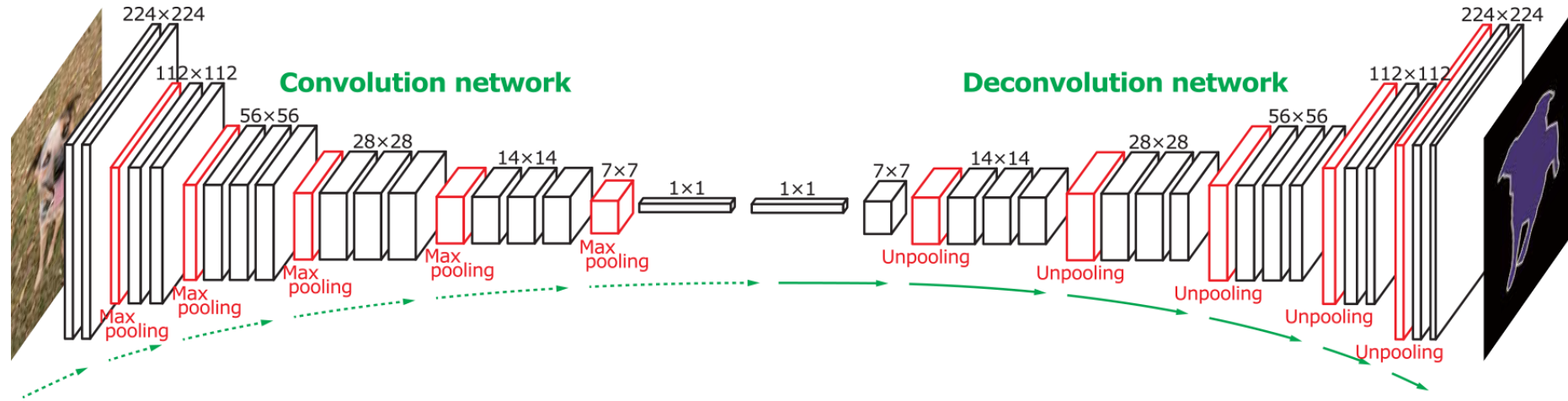


# Divers - Deconvnet



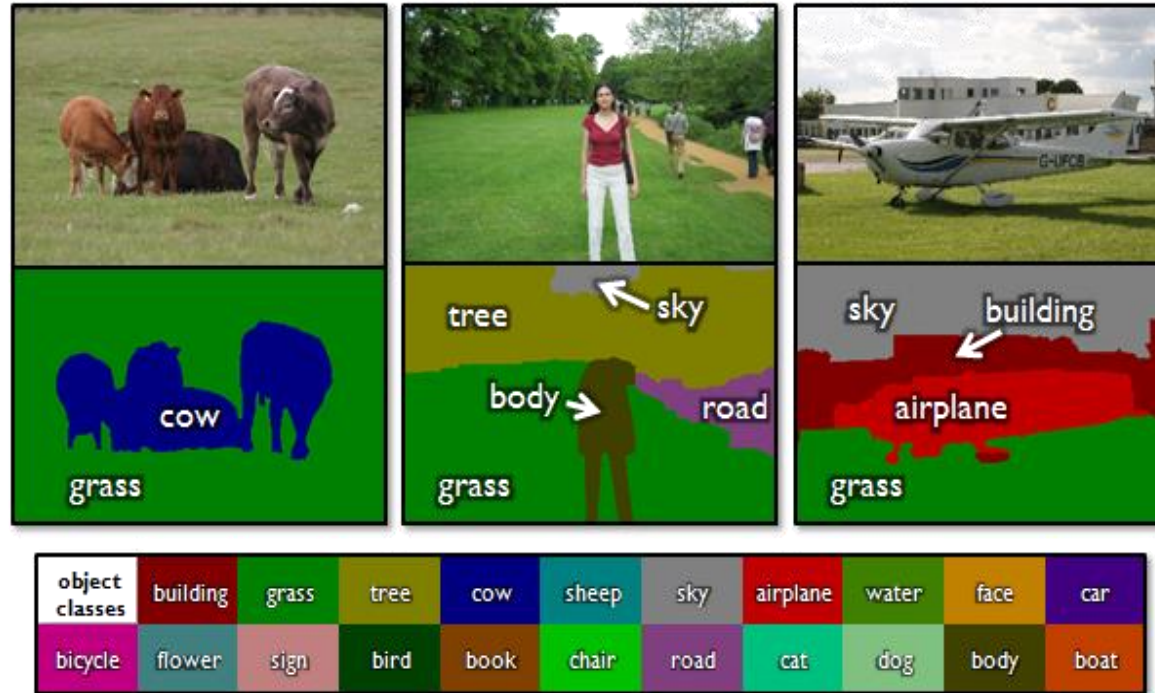
Source : <http://arxiv.org/abs/1311.2901>

# Divers - Deconvnet



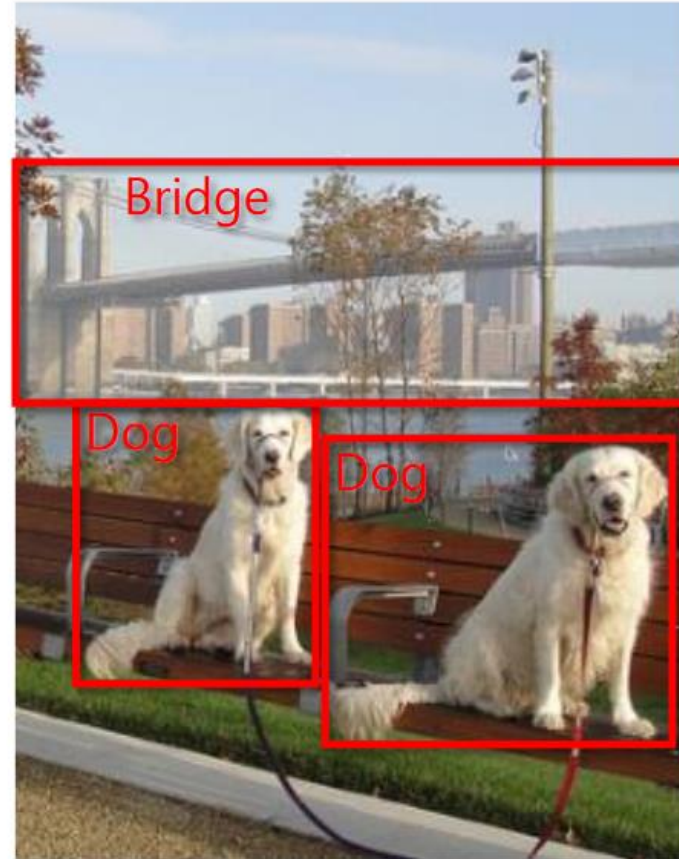
<https://www.youtube.com/watch?v=AgkflQ4IGaM>

# Divers - segmentation



[Blog sur le sujet](#)

# Divers - détection



Source : [Detection tutorial for caffe](#)

# Divers - conseils d'architecture et autre

- Aucun
- <http://cs231n.github.io/convolutional-networks/>

# Ressources

- [What makes a good selfie?](#)
- [CNN for text classification - MxNet](#)
- [CNN for translation – Torch](#)
- [Deep Visual .... Generating Image Descriptions](#)

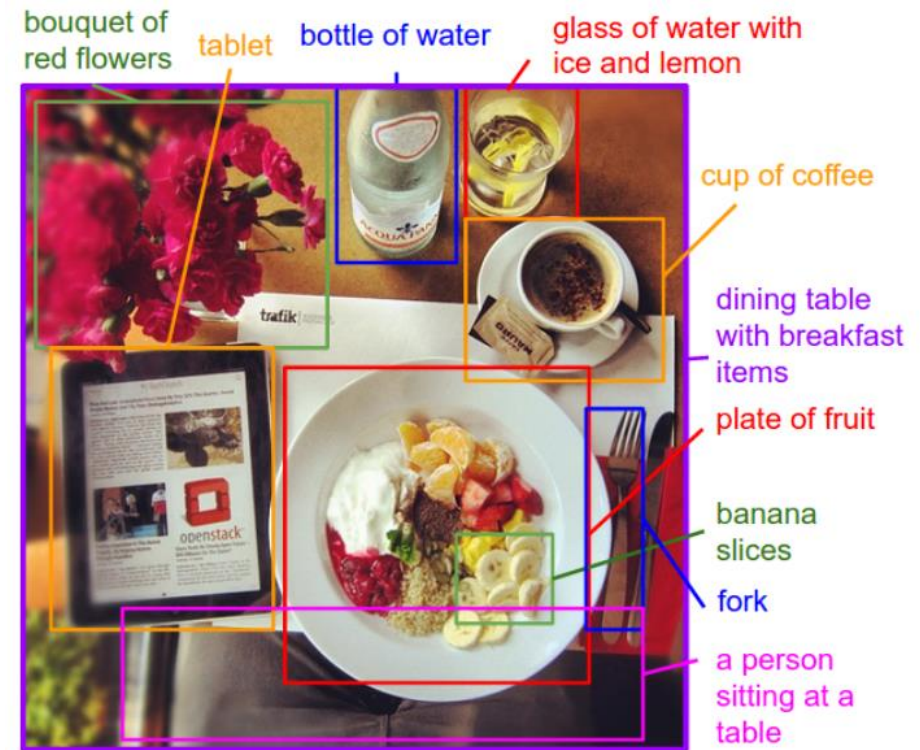
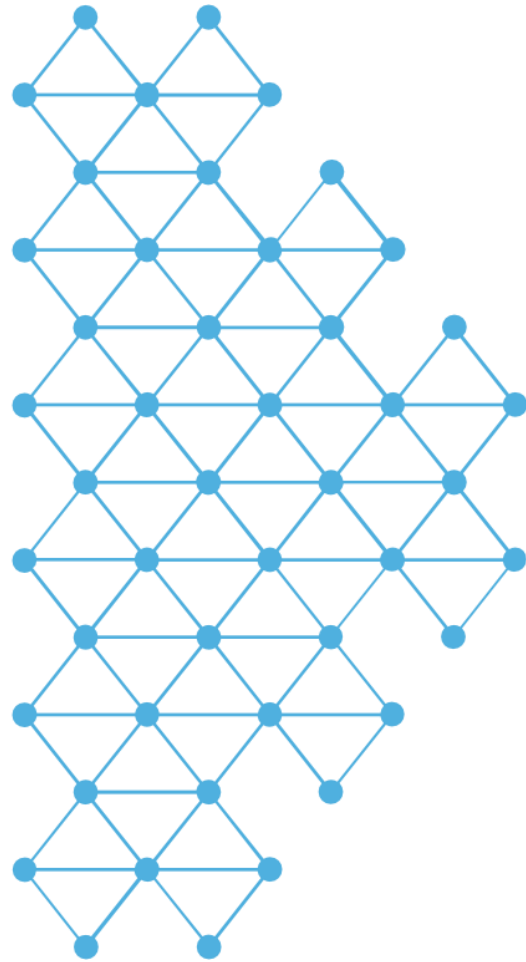
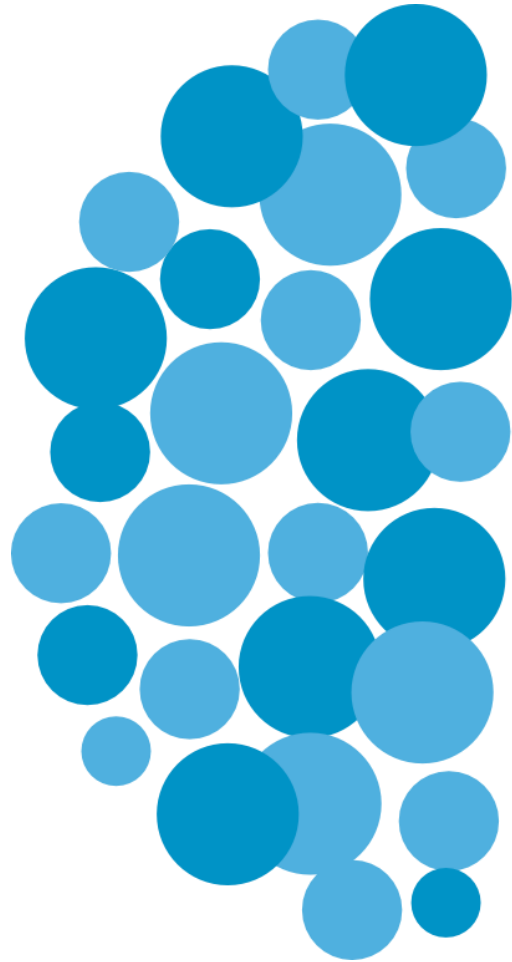


Figure 1. Motivation/Concept Figure: Our model treats language as a rich label space and generates descriptions of image regions.



# Questions





MILA