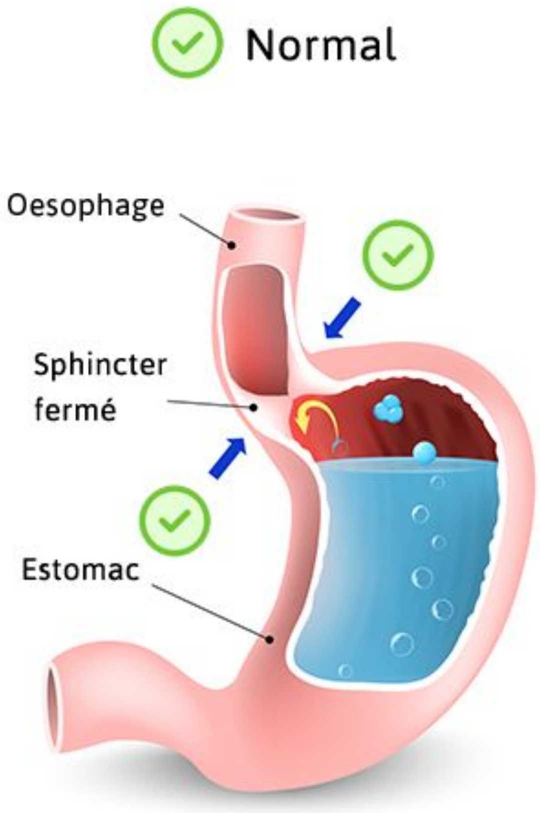
Challenge Data ENS 2019 :

“Screening and Diagnosis of esophageal cancer from in-vivo microscopy images”

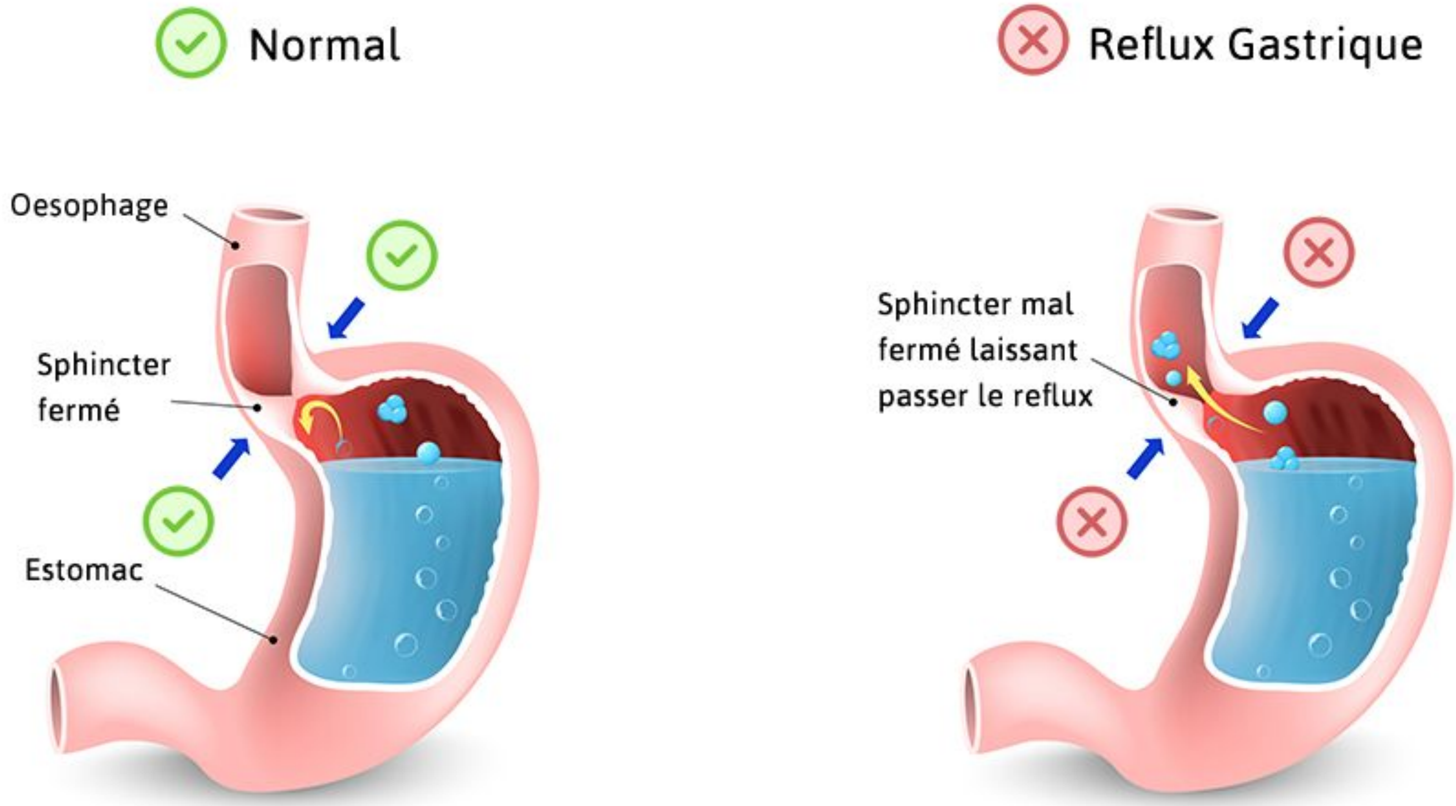
proposé par : Mauna Kea Technologies

Jonas Maison

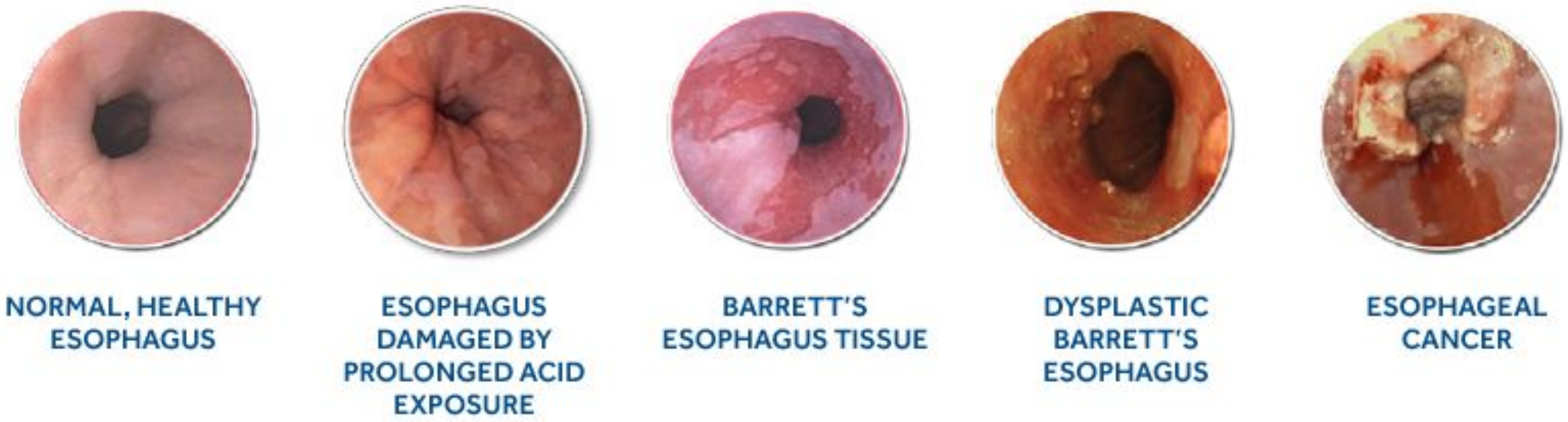
Présentation du challenge



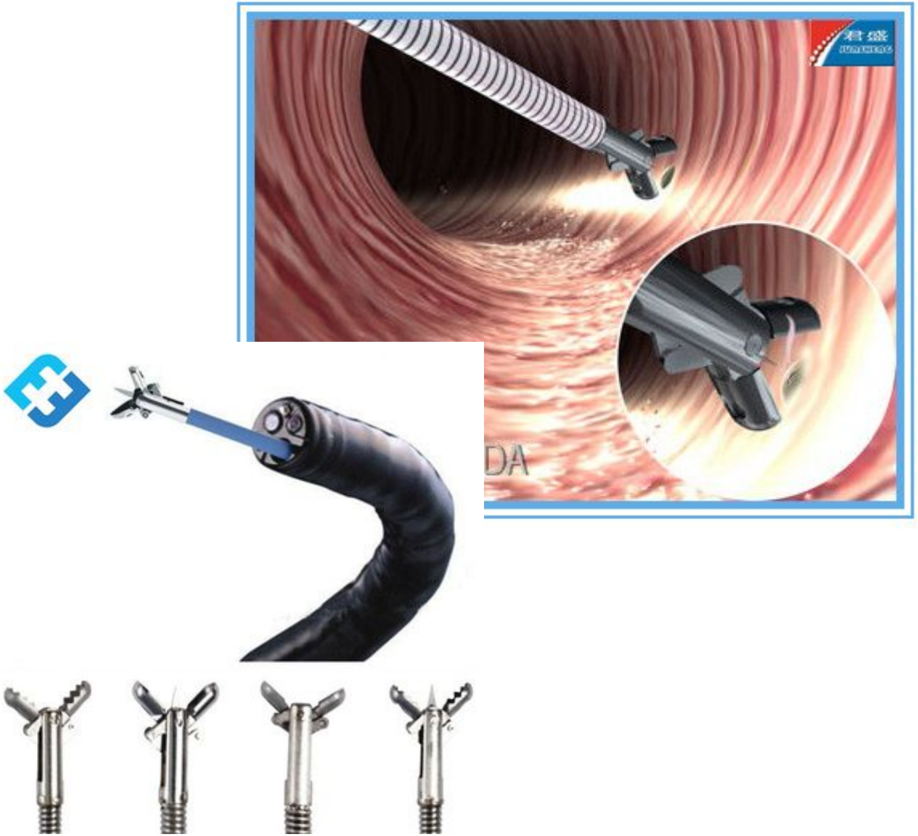
Présentation du challenge

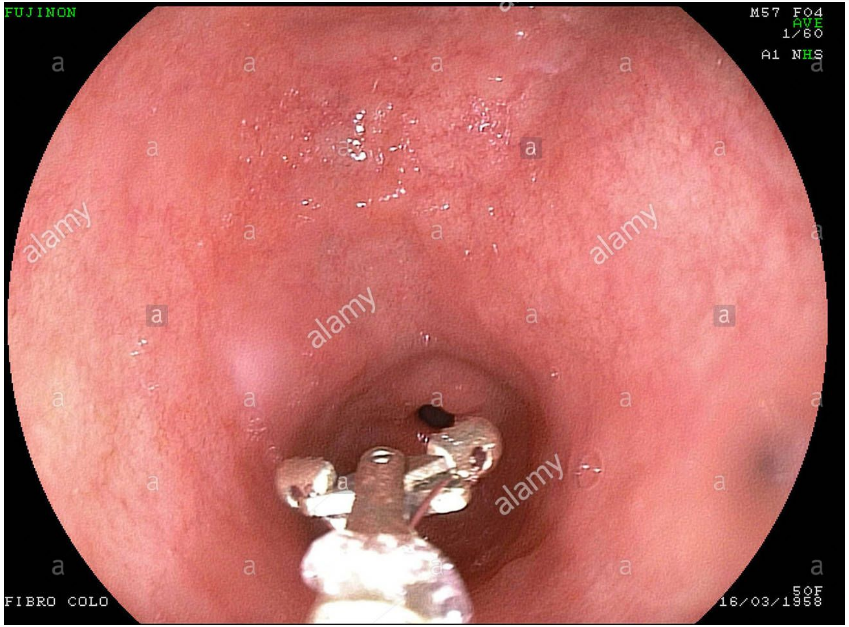


Présentation du challenge

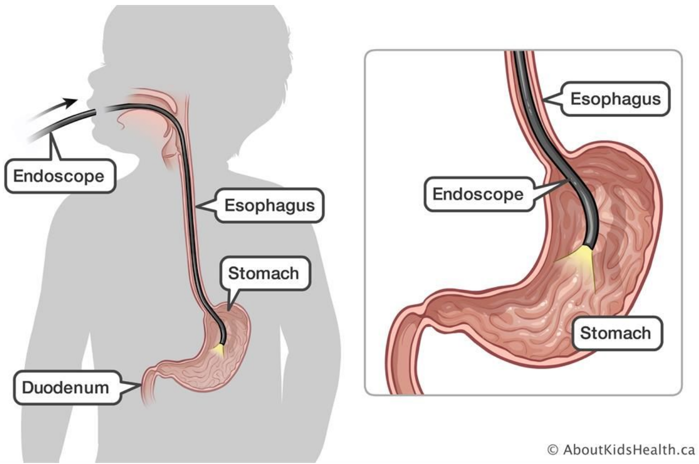


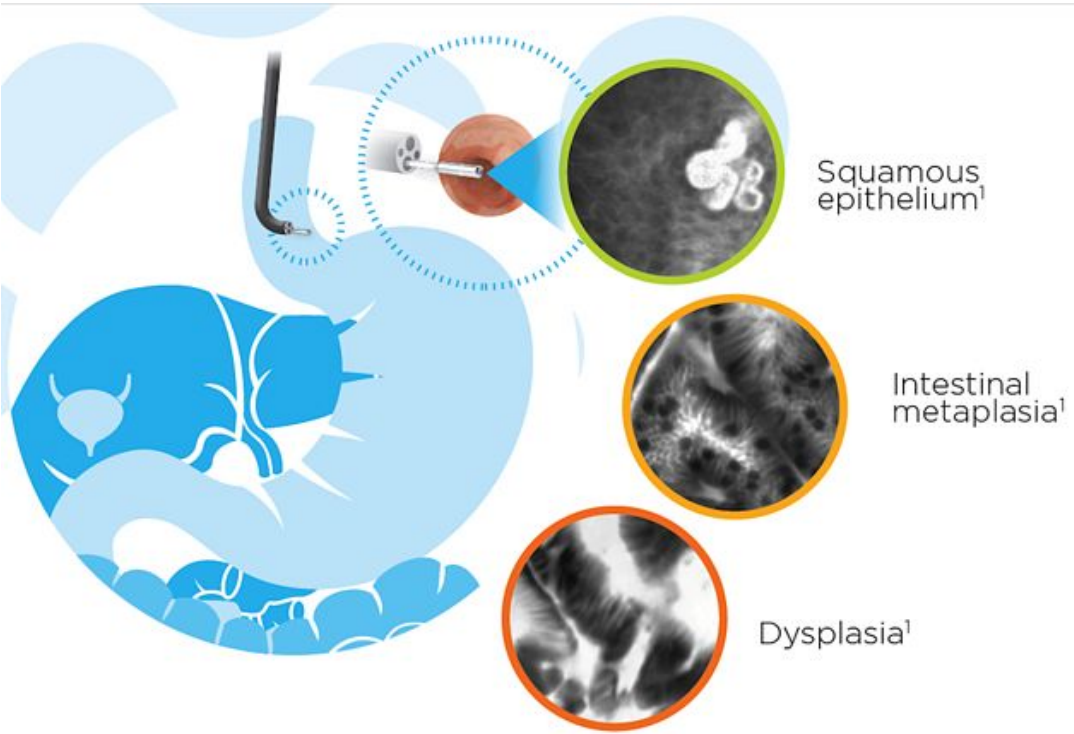
Présentation du challenge





Présentation du challenge





Challenge

- Problème de classification d'images
- Training set : 9446 images de 44 patients
- 4 classes :
  - Squamous Epithelium : 1469 images
  - Intestinal Metaplasia : 3177 images
  - Gastric Metaplasia : 1206 images
  - Dysplasia/Cancer : 3594 images
- Set de test (sans labels) : 1715 images de 17 patients
- Pas d'overlap entre train/test sets (par rapport aux patients)
- Métrique : non-weighted multiclass accuracy
- Baseline : 78% (CNN simple à 3 couches)
- Objectif de 99% pour que le produit soit commercialisable

Challenge

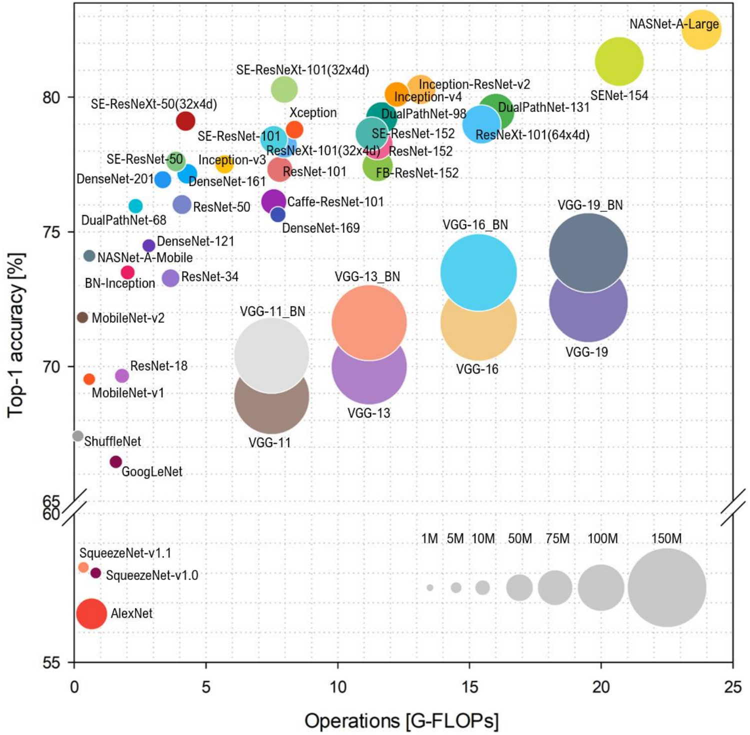
- Peu de publications dans la littérature sur ce sujet
- Approche choisie : Deep learning avec CNN

Choix de l’architecture

Benchmark Analysis of Representative Deep Neural Network Architectures

Simone Bianco, Remi Cadene, Luigi Celona, Paolo Napoletano

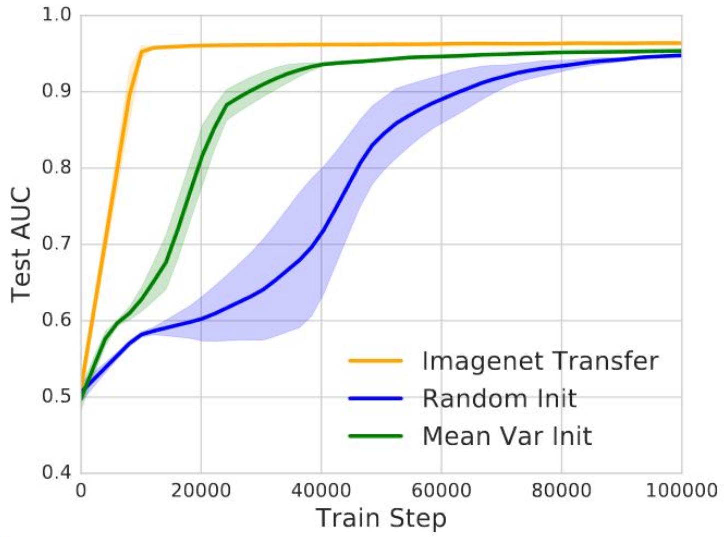
- Architectures testées :
    - DenseNet
    - SE-ResNext101
    - SE-ResNet152
    - InceptionResNetV2



Initialisation des poids

En général, 3 méthodes :

- Random
- Transfer learning sur ImageNet
- Auto-encodeur



How transferable are features in deep neural networks?

Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson

Transfusion: Understanding Transfer Learning with Applications to Medical Imaging

Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, Samy Bengio

Taille de batch

Grande taille de batch :

Augment your batch: better training with larger batches

Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler, Daniel Soudry

Taille de batch

Grande taille de batch :

Augment your batch: better training with larger batches
Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler, Daniel Soudry

Petite taille de batch :

Revisiting Small Batch Training for Deep Neural Networks
Dominic Masters, Carlo Luschi

On the Computational Inefficiency of Large Batch Sizes for Stochastic Gradient Descent
Noah Golmant, Nikita Vemuri, Zhewei Yao, Vladimir Feinberg, Amir Gholami, Kai Rothauge, Michael W. Mahoney, Joseph Gonzalez

On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima
Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, Ping Tak Peter Tang

A Walk with SGD
Chen Xing, Devansh Arpit, Christos Tsirigotis, Yoshua Bengio

Taille de batch

Grande taille de batch :

Augment your batch: better training with larger batches
Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler, Daniel Soudry

Petite taille de batch :

Revisiting Small Batch Training for Deep Neural Networks
Dominic Masters, Carlo Luschi

On the Computational Inefficiency of Large Batch Sizes for Stochastic Gradient Descent
Noah Golmant, Nikita Vemuri, Zhewei Yao, Vladimir Feinberg, Amir Gholami, Kai Rothauge, Michael W. Mahoney, Joseph Gonzalez

On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima
Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, Ping Tak Peter Tang

A Walk with SGD
Chen Xing, Devansh Arpit, Christos Tsirigotis, Yoshua Bengio

Les deux :

Don't Decay the Learning Rate, Increase the Batch Size
Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, Quoc V. Le

AdaBatch: Adaptive Batch Sizes for Training Deep Neural Networks
Aditya Devarakonda, Maxim Naumov, Michael Garland

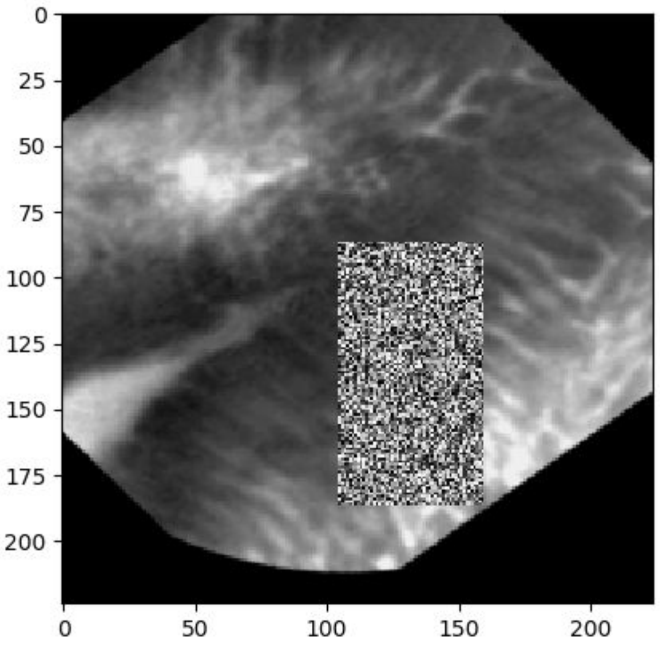
Coupling Adaptive Batch Sizes with Learning Rates
Lukas Balles, Javier Romero, Philipp Hennig

Augmentation d’image

- Permet d’éviter l’overfitting, améliore le score
- Transformations (aléatoires) :
  - Rotation 360°
  - Reshape (224, 224) ou (299, 299)
  - Déformations élastiques
  - Translation
  - Flips horizontaux et verticaux
  - Zoom (in and out)
  - Color Jitter (luminosité, contraste, …)
  - RandomErase

Random Erasing Data Augmentation

Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, Yi Yang



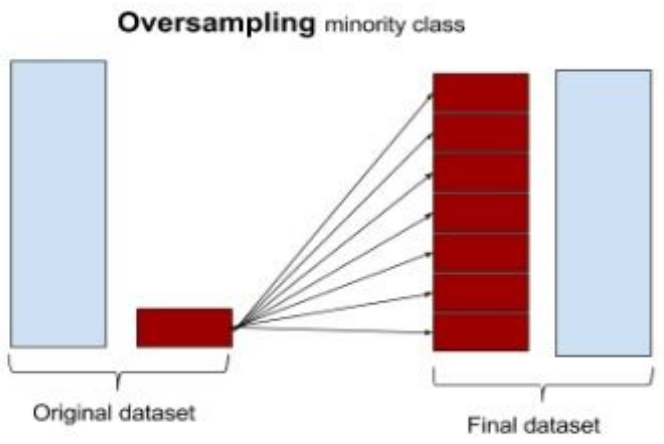
Équilibrage des classes

A systematic study of the class imbalance problem in convolutional neural networks

Mateusz Buda, Atsuto Maki, Maciej A. Mazurowski

Oversampling :

- Simple à utiliser
- Marche bien avec l'augmentation d'image



Fonction de coût

- Softmax + Cross entropy
- Non-weighted durant l’entraînement (car oversampling)
- Weighted durant l’évaluation

SGD ou autre ?

The Marginal Value of Adaptive Gradient Methods in Machine Learning

Ashia C. Wilson, Rebecca Roelofs, Mitchell Stern, Nathan Srebro, Benjamin Recht

Improving Generalization Performance by Switching from Adam to SGD

Nitish Shirish Keskar, Richard Socher

⇒ **SGD** (5e-3 learning rate, 0.9 momentum, 1e-4 weight decay, ...)

Optimiseur

SGD ou autre ?

The Marginal Value of Adaptive Gradient Methods in Machine Learning
Ashia C. Wilson, Rebecca Roelofs, Mitchell Stern, Nathan Srebro, Benjamin Recht

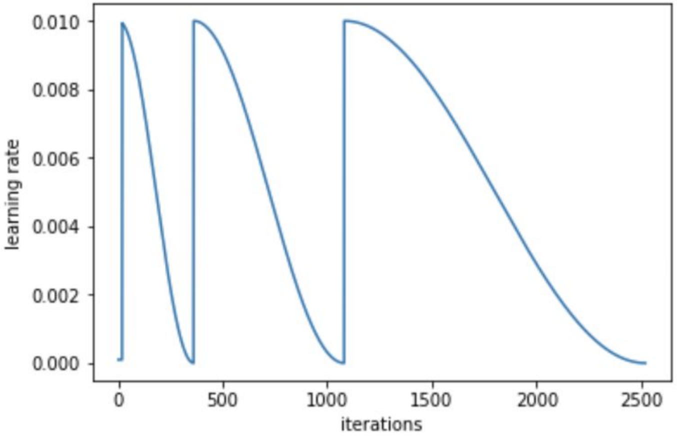
Improving Generalization Performance by Switching from Adam to SGD
Nitish Shirish Keskar, Richard Socher

⇒ **SGD** (5e-3 learning rate, 0.9 momentum, 1e-4 weight decay, ...)

Learning rate schedule :

SGDR: Stochastic Gradient Descent with Warm Restarts
Ilya Loshchilov, Frank Hutter

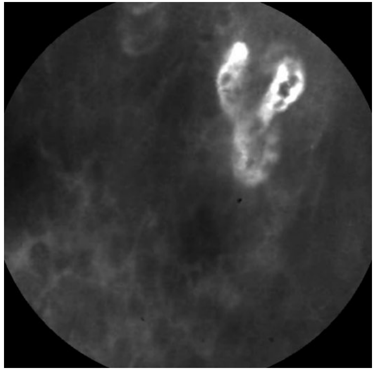
Cyclical Learning Rates for Training Neural Networks
Leslie N. Smith

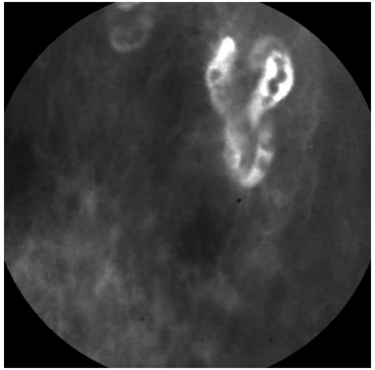


Partitionnement des sets

Constats :

- Les classes sont mal équilibrées
- Forte corrélation entre les images d'un même patient





Solution choisie :

- Partitionnement aléatoire :
  - les images sont très variables d'un patient à l'autre
  - l'augmentation d'image permet de limiter l'overlap entre training set et test set

- Pas de K-fold : trop long à entraîner (5 modèles par architecture)

- Simple train (90%) / test (10%) ⇒ 1 modèle par architecture

Résultats

Solution	Score
DenseNet210	0.887
DenseNet169	0.890
SE-ResNext101	0.892
Inception ResNet V2	0.893
DenseNet169	0.895
Ensemble (3 modèles)	0.914
Ensemble (5 modèles)	**0.918**

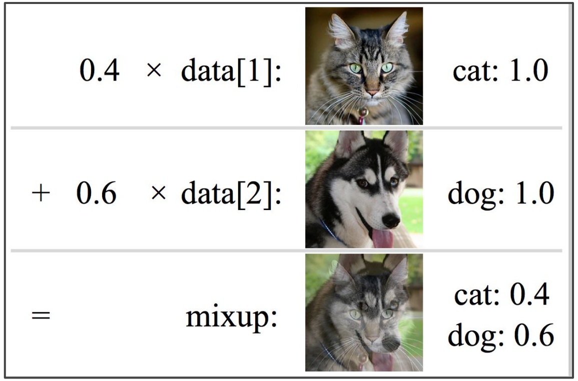
Améliorations

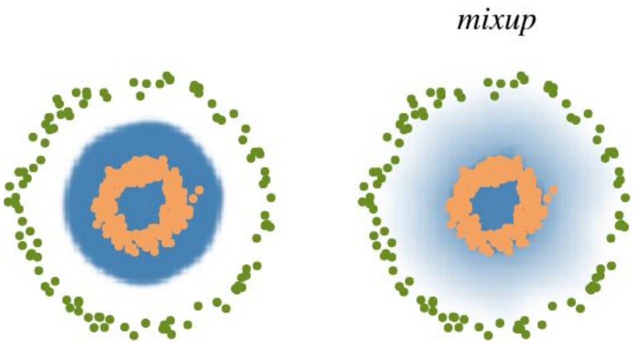
Les réseaux sont “trop sûrs” de leurs prédictions (même quand ils se trompent)

Améliorations

Les réseaux sont “trop sûrs” de leurs prédictions (même quand ils se trompent)

⇒ MIXUP





Permet d’améliorer l’ensemble en prenant en compte les probabilités

Manifold Mixup: Better Representations by Interpolating Hidden States
Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, Aaron Courville, David Lopez-Paz, Yoshua Bengio

mixup: Beyond Empirical Risk Minimization
Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz

Between-class Learning for Image Classification
Yuji Tokozume, Yoshitaka Ushiku, Tatsuya Harada

Deux constats :

1) 1715 images du set de test non exploitées
2) Les images d'un même patient appartiennent probablement à la même classe

⇒ Pseudo-labeling

Score final

Ranking	Date	User(s)	Public score
1	March 4, 2019, 4:51 p.m.	lina.mezghani & jbsevestre	0.9732
2	March 6, 2019, 11:06 a.m.	Louis_Verret	0.9662
3	Feb. 17, 2019, 4:54 a.m.	tutti_frutti	0.9592
4	March 14, 2019, 5 p.m.	jonas1312	0.9487
5	March 15, 2019, 1:38 p.m.	bplaterrier	0.9033
6	Feb. 14, 2019, 11:39 p.m.	louist	0.8904
7	Feb. 12, 2019, 5:50 a.m.	axeldldl	0.8834
8	Feb. 21, 2019, 11:24 a.m.	Cantor	0.8800
9	March 5, 2019, 8:22 p.m.	PeterA2Z	0.8660
10	Feb. 27, 2019, 10:06 p.m.	Yoann	0.8566
11	Feb. 12, 2019, 8:45 a.m.	Deleted user	0.8357
12	-	benchmark	0.7844
13	Feb. 22, 2019, 4:11 p.m.	marczakl	0.7576
14	Feb. 12, 2019, 12:49 p.m.	umpalumpa	0.7494
15	March 2, 2019, 3:10 p.m.	Abribus37	0.6923
16	Feb. 12, 2019, 11:31 a.m.	Adeikalam	0.5746
17	Jan. 27, 2019, 8:11 p.m.	johny	0.3380
18	Feb. 8, 2019, 8:11 a.m.	edenpowa	0.3263
19	March 4, 2019, 10:06 p.m.	moai	0.2657

Outils utilisés

Modèles pré-entraînés : https://github.com/Cadene/pretrained-models.pytorch

Bibliographie : https://arxiv.org/ et http://www.arxiv-sanity.com/